

December 2013

# Civil Good - A Platform For Sustainable and Inclusive Online Discussion

David C. Pounds

*Worcester Polytechnic Institute*

Ian Adam Naval

*Worcester Polytechnic Institute*

John Duc Pham

*Worcester Polytechnic Institute*

Michael A. Perrone

*Worcester Polytechnic Institute*

Sophia Elizabeth Pounds

*Worcester Polytechnic Institute*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.wpi.edu/iqp-all>

---

## Repository Citation

Pounds, D. C., Naval, I. A., Pham, J. D., Perrone, M. A., Pounds, S. E., Malis, S. M., O'Connor, T. J., & Narayan, T. (2013). *Civil Good - A Platform For Sustainable and Inclusive Online Discussion*. Retrieved from <https://digitalcommons.wpi.edu/iqp-all/264>

This Unrestricted is brought to you for free and open access by the Interactive Qualifying Projects at Digital WPI. It has been accepted for inclusion in Interactive Qualifying Projects (All Years) by an authorized administrator of Digital WPI. For more information, please contact [digitalwpi@wpi.edu](mailto:digitalwpi@wpi.edu).

---

**Author**

David C. Pounds, Ian Adam Naval, John Duc Pham, Michael A. Perrone, Sophia Elizabeth Pounds, Steven Mark Malis, Thomas John O'Connor, and Tushar Narayan



# WPI

## **Civil Good:**

A Platform For Sustainable and Inclusive Online Discussion

An Interactive Qualifying Project submitted to the faculty of  
Worcester Polytechnic Institute

In partial fulfillment of the requirements for the degree of Bachelor of Science  
by:

Steven Malis (Computer Science),  
Tushar Narayan (Computer Science),  
Ian Naval (Computer Science),  
Thomas O'Connor (Biochemistry),  
Michael Perrone (Physics and Mathematics),  
John Pham (Computer Science),  
David Pounds (Computer Science and Robotics Engineering),

December 19, 2013

Submitted to:  
Professor Craig Shue, WPI Advisor  
Alan Mandel, Creator of the Civil Good concept

# Contents

<b>1</b>	<b>Executive Summary</b>	<b>1</b>
1.1	Overview of Recommendations . . . . .	2
<b>2</b>	<b>Authorship</b>	<b>5</b>
<b>3</b>	<b>Introduction</b>	<b>10</b>
3.1	Existing Work - Similar Websites . . . . .	11
<b>4</b>	<b>Psychology</b>	<b>17</b>
4.1	Online Disinhibition . . . . .	17
4.2	Format of Discussions . . . . .	22
4.3	Reducing Bias with Self-Affirmation . . . . .	28
4.4	Other Psychological Influences . . . . .	34
<b>5</b>	<b>Legal Issues</b>	<b>38</b>
5.1	Personally Identifiable Information . . . . .	38
5.2	Intellectual Property . . . . .	42
5.3	Defamation . . . . .	45
5.4	Information Requests . . . . .	46
5.5	Use by Minors . . . . .	49
5.6	General Litigation Avoidance and Defense . . . . .	51
<b>6</b>	<b>Societal Impact</b>	<b>52</b>
6.1	Political Polarization . . . . .	52
6.2	Minority Opinion Representation . . . . .	55
6.3	History and Political Climate . . . . .	59
6.4	TED Talks Model . . . . .	64

<b>7</b>	<b>Infrastructure</b>	<b>66</b>
7.1	Web Hosting Requirements . . . . .	66
7.2	Databases . . . . .	71
7.3	Programming Languages and Open Source Technology . . . . .	77
<b>8</b>	<b>Incremental Refinement and Data-Driven Development</b>	<b>86</b>
8.1	Background . . . . .	86
8.2	Incremental Refinement . . . . .	93
8.3	Data-Driven Development . . . . .	94
8.4	Conclusion . . . . .	99
<b>9</b>	<b>Security</b>	<b>100</b>
9.1	Database Security . . . . .	100
9.2	Common Attacks . . . . .	103
9.3	Encryption . . . . .	111
9.4	Password Security . . . . .	115
<b>10</b>	<b>Privacy</b>	<b>122</b>
10.1	Minimal Tracking Requirements . . . . .	122
10.2	Browser Fingerprinting . . . . .	129
10.3	Third Party Privacy Risks . . . . .	132
10.4	Privacy Policy . . . . .	137
<b>11</b>	<b>Community - Developers and Users</b>	<b>141</b>
11.1	Attracting Developers . . . . .	141
11.2	User Trends and Retention . . . . .	146
<b>12</b>	<b>Business Model</b>	<b>152</b>
12.1	Operational Costs . . . . .	152

12.2 501(c)(3) Status . . . . .	155
12.3 Marketing Costs . . . . .	160
12.4 Revenue . . . . .	166
<b>13 User Interaction</b>	<b>175</b>
13.1 Related Work . . . . .	175
13.2 Recommended Changes to Website Design . . . . .	180
13.3 Conversation Page Improvements . . . . .	192
13.4 Rating System . . . . .	199
13.5 Experiment . . . . .	203
<b>14 Mobile Presence</b>	<b>214</b>
14.1 Mobile Demographic . . . . .	214
14.2 Mobile Application Security . . . . .	217
14.3 Mobile Web Design . . . . .	221
14.4 Mobile Applications vs. Websites . . . . .	223
<b>15 Conclusions</b>	<b>229</b>
<b>16 Bibliography</b>	<b>244</b>
<b>17 Acknowledgements</b>	<b>297</b>
<b>Appendices</b>	<b>298</b>
<b>A User Flow</b>	<b>298</b>
<b>B Interview Transcript, Mr. Eugene Suzuki</b>	<b>301</b>
<b>C Interview Transcript, Professor Soussan Djamasbi</b>	<b>316</b>

D Interview Transcript, Mr. Rob Bertsche and Mr. Peter Caruso	330
E Privacy Policy	342
F Website Mockups	344

## List of Figures

1	High-End Cost Per Click (CPC) Details . . . . .	162
2	High-End Recommended Bid Data . . . . .	162
3	Low-End Recommended Bid Data . . . . .	162
4	Experiment data—Hours spent on the Internet in an average week . .	207

## List of Tables

1	Ringlemann's Results . . . . .	24
2	Independent t-Tests for Differences in Perceptions of Participants in One-on-One Group Interactions. . . . .	25
3	Traffic Data . . . . .	153
4	Web Hosting Prices . . . . .	155
5	Results from Google Keyword Planner . . . . .	161
6	Google Display Planner Data . . . . .	163
7	Rating Criteria . . . . .	202
8	Experiment Data—Number of participants self-identifying at a par- ticular level of computer skill . . . . .	207
9	Responses: Original Design . . . . .	209
10	Responses: Modified Design . . . . .	209
11	Experiment Data—Average ranking and standard deviations for each question for the two website versions . . . . .	210



## Abstract

Civil Good is a website concept proposed by Alan Mandel with the goal of enabling safe, anonymous, productive, and civil discourse without the disruptive behavior and language common to much of the Internet. The goal of Civil Good is to improve the critical thinking and discussion skills of its users while combating the effects of political polarization and misinformation in society.

This paper analyzes Mandel’s proposed concept, providing additional research to either support or refute the various features proposed, and recommendations to simplify user interactions. It also examines topics mentioned only briefly or not discussed by Mandel, such as data protection methods, the psychology of Web browsing, marketing, operational costs, legal issues, monetization options, and mobile presence.

## 1 Executive Summary

Discussion on the Internet generally tends to conform to the “lowest-common-denominator” [86]. The discussion in virtual chat rooms often tends to veer towards mundane topics, or devolves into the participants venting their frustration with a specific societal development. Serious conversations get lost in the multitude of less serious comments. With this aspect of online discussions permeating every popular website, there does not exist an online forum that meets the needs of people who would like a civil online environment to discuss and debate the important topics of the day with each other.

Civil Good is a website concept proposed by Alan Mandel. The website is aimed at providing a forum for anonymous and civil debate on a wide variety of topics, and is primarily oriented to the discussion of topics situated at the intersection

of society, politics, and economics [252].

The act of expressing deeply held beliefs about the day’s pressing issues brings with it the associated risks of alienation, the potential for judgment, and the lack of privacy. Further, when talking offline, humans tend to talk to others who share similar views, and thus the potential for thought-provoking debate is greatly reduced [252]. By transferring those conversations online, making them anonymous, and enabling interactions with strangers hundreds—if not millions—of miles away, Civil Good would provide a service capable of fostering discussion, increasing communication across geopolitical boundaries, and ensuring a safe environment for people to express their true beliefs.

However, those benefits do also raise questions. How does the website ensure respectful discussion while also honoring individual privacy? How does Civil Good monetize the website enough to be self-sustaining while also making sure that user information is not inadvertently shared with advertisers? How does the website protect the data in a manner that minimizes the opportunities for a governmental organization to either demand or steal the data and potentially identify (and subsequently persecute) a specific user?

These questions—and others—need to be resolved if the idea of Civil Good is to achieve its goal of significant societal impact. Along with these questions, in this project we sought to help Mandel further develop his ideas by providing research to either support or refute Mandel’s thoughts on the open questions, and to provide depth on the topic areas as yet unexplored.

## **1.1 Overview of Recommendations**

We examined the psychology of human communication, both offline and online, and the benefits of group discussion vis-à-vis one-on-one conversation. We found that

one-on-one conversations are superior to group discussions, as far as the quality of the conversation is concerned.

We also researched the bigger societal impact that a forum like Civil Good could have. We determined how it could be a force for reducing the polarization that is currently rampant in politics, and how it could improve the representation that the minority currently has in voicing their opinion on the big-ticket issues of the day.

We also researched, developed, and proposed a strategy that would ensure that the website can become self-sufficient after accounting for the developmental, system and bandwidth costs. While there are many variables involved with the actual costs associated with the website, we were able to propose various scenarios and offer a monetization strategy for each of them.

We also researched the various types of informational databases with an emphasis on the related security concerns, and made a recommendation on the most applicable type of database for the website. Since preserving user anonymity lies at the core of the idea of Civil Good, we researched the various types of attacks and security loopholes that affect websites. We made recommendations to ensure that Civil Good incorporates security as a core developmental goal so as to be able to thwart attacks against the website.

We explored methods for improving Civil Good once the website is running. We found that incremental refinement of website design in concert with data driven design is the best approach. We recommend that aggregated and anonymized data be collected from the interactions that users have with Civil Good to further shape and develop the website.

We researched how to foster a positive environment on the website by exploring human psychological biases and how they can be combated. We recommended

ways of integrating existing psychological research into the design of the website. We also discussed the effects of user awareness of psychological effects on user behavior.

We proposed an initial privacy policy that was reviewed by an attorney at law. We also investigated the legal issues that United States-based social media websites face, and recommended policies for dealing with these issues. We also recommended that Civil Good seek legal counsel so as to keep apprised of any new legal developments.

We also researched the option of developing a companion mobile website or application for Civil Good. After finding that a mobile website was the better option, we developed a list of the changes that the mobile website would have with respect to the desktop version.

Through the process of a literature review, we also developed a set of recommended changes to the design of the website proposed by Mandel. We conducted one week of experimental testing using mock ups of Mandel's original design and a design that incorporated our recommended changes to gather data on which version was more user friendly. We were able to draw clear trends from the data that indicated that the changes we proposed did in fact increase the user friendliness of the website.

## 2 Authorship

Every team member helped by providing feedback during the editing process.

### Introduction

**Executive Summary, Introduction, Section Introductions:** Written by Tushar Narayan; peer reviewed and edited by all team members.

**Existing Work - Similar Websites:** Written by Michael Perrone; peer reviewed by Tushar Narayan; edited by Steven Malis.

### Psychology

**Online Disinhibition:** Written by John Pham; peer reviewed by Ian Naval and David Pounds; edited by Tushar Narayan.

**Format of Discussions:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan and David Pounds; edited by John Pham.

**Reducing Bias with Self-Affirmation:** Written by John Pham; peer reviewed by David Pounds and Tushar Narayan; edited by Steven Malis.

**Other Psychological Influences:** Written by John Pham; peer reviewed by Tushar Narayan; edited by Steven Malis.

### Legal Issues

**Personally Identifiable Information:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

**Intellectual Property:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

**Defamation:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

**Information Requests:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

**Use by Minors:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

**General Litigation Avoidance and Defense:** Written by Steven Malis; peer reviewed and edited by Ian Naval.

## **Societal Impact**

**Political Polarization:** Written by John Pham; peer reviewed by David Pounds and Michael Perrone; edited by Tushar Narayan.

**Minority Opinion Representation:** Written by John Pham; peer reviewed by Tushar Narayan and Steven Malis; edited by David Pounds.

**History and Political Climate:** Written by John Pham; peer reviewed by Tushar Narayan and Steven Malis; edited by Ian Naval.

**TED Talks Model:** Written by Michael Perrone; peer reviewed by Thomas O'Connor and David Pounds; edited by John Pham.

## **Infrastructure**

**Web Hosting Requirements:** Written by David Pounds; peer reviewed by Tushar Narayan; edited by Steven Malis.

**Databases:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan and Steven Malis; edited by Ian Naval.

**Programming Languages and Open-Source Technology:** Written by Thomas O'Connor; peer reviewed by Michael Perrone and John Pham; edited by David Pounds and Ian Naval.

## **Incremental Refinement and Data-Driven Development**

**Background:** Written by Michael Perrone; peer reviewed by Thomas O'Connor, John Pham, Ian Naval, Steven Malis, David Pounds and Tushar Narayan.

**Incremental Refinement:** Written by Michael Perrone; peer reviewed by Thomas O'Connor, John Pham, Ian Naval, Steven Malis, David Pounds and Tushar Narayan.

**Data-Driven Development:** Written by Michael Perrone; peer reviewed by Thomas O'Connor, John Pham, Ian Naval, Steven Malis, David Pounds and Tushar Narayan.

**Conclusion:** Written by Michael Perrone; peer reviewed by Thomas O'Connor, John Pham, Ian Naval, Steven Malis, David Pounds and Tushar Narayan.

## **Security**

**Database Security:** Written by Ian Naval; peer reviewed by Steven Malis; edited by David Pounds.

**Common Attacks:** Written by Ian Naval; peer reviewed by Tushar Narayan and Steven Malis; edited by Thomas O'Connor.

**Encryption:** Written by Ian Naval; peer reviewed by David Pounds and Tushar Narayan; edited by David Pounds.

**Password Security:** Written by Ian Naval; peer reviewed by Steven Malis and David Pounds; edited by Thomas O'Connor.

## **Privacy**

**Browser Fingerprinting:** Written by Ian Naval; peer reviewed by David Pounds; edited by Tushar Narayan.

**Third Party Privacy Risks:** Written by David Pounds; peer reviewed by John Pham; edited by Tushar Narayan.

**Minimal Tracking Requirements:** Written by David Pounds; peer reviewed by Tushar Narayan; edited by John Pham.

**Privacy Policy:** Written by Tushar Narayan; peer reviewed by Steven Malis and David Pounds; edited by Thomas O'Connor.

## **Community - Developers and Users**

**Attracting Developers:** Written by John Pham; peer reviewed by Steven Malis and Michael Perrone; edited by Tushar Narayan.

**User Trends and Retention:** Written by John Pham; peer reviewed by Tushar Narayan; edited by Thomas O'Connor.

## **Business Model**

**Operational Costs:** Written by David Pounds; peer reviewed by Michael Perrone; edited by Ian Naval.

**501(c)(3) Status:** Written by David Pounds; peer reviewed by Tushar Narayan; edited by John Pham.

**Marketing Costs:** Written by Ian Naval; peer reviewed by David Pounds and Michael Perrone; edited by Tushar Narayan.

**Revenue:** Written by Ian Naval; peer reviewed by David Pounds and Tushar Narayan; edited by John Pham.



## User Interaction

**Related Work, Recommended Changes to Website Design:** Written by Steven Malis; peer reviewed by Tushar Narayan and Michael Perrone; edited by David Pounds.

**Conversation Page Improvements:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan; edited by John Pham.

**Rating System:** Written by David Pounds; peer reviewed by Tushar Narayan; edited by Steven Malis.

**Experiment:** Written by Tushar Narayan; peer reviewed by Thomas O'Connor and John Pham; edited by David Pounds.

## Mobile Presence

**Mobile Demographic:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan; edited by Ian Naval.

**Mobile Application Security:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan and Ian Naval; edited by David Pounds.

**Mobile Web Design:** Written by Thomas O'Connor; peer reviewed and edited by Steven Malis.

**Mobile Applications vs. Websites:** Written by Thomas O'Connor; peer reviewed by Tushar Narayan; edited by Ian Naval.

The **Conclusions** section was written and edited collaboratively by every member of the team.

### 3 Introduction

Discussion on the Internet is currently flawed. While the very nature of the Internet lends itself to numerous websites and other communication methods that support the technical aspects of conducting an online conversation, the general tendency of an online conversation is to descend into a “lowest-common-denominator exchange” at best and a brawl of swear words and name-calling at worst [86]. This causes the exchange to stray from the original objective of having a conversation conducive to respectful debate and an honest trade of ideas.

This scenario lends itself to disruption. The force of the Internet as an engine for starting and accelerating public discussion is undeniable [304]. The very fact that people can transcend genders, cultures, geographic boundaries, great distances, and tyrannical regimes to come together in a medium such as the Internet to exchange their ideas, opinions, and feelings is powerful. A forum that allows true freedom of speech—free from the fear of oppression—is required to spark civil discourse on a scale never before seen.

The “Civil Good” idea, created by Alan Mandel, is a website aimed at fostering online communication in a manner that is constructive and conducive to coherently expressing opinions. The core of the idea is total privacy and anonymity, guaranteeing immunity from prosecution, open discussions with a conversational partner meeting a highly specific set of demographics, and an environment that is open to expressing deeply held opinions while fostering constructive debate.

This project examines the various major facets of the “Civil Good” website, delving into issues that impact conversations both online and offline. The goal of this project is to add more detail and explore avenues that would push this idea towards becoming both a reality and a force for social change.

### 3.1 Existing Work - Similar Websites

Although Civil Good is not the first site to encourage debate on the Web, it is the first site that meets all the criteria of the Civil Good project [252]. When designing a site like Civil Good, it is important to consider what has already been done in similar fields to better inform design decisions and develop bases for comparison.

One of the underlying principles of Civil Good is that people cannot possibly debate in a meaningful way if they do not have the relevant facts. Civil Good addresses this by providing statistics and information relevant to a conversation topic before that conversation begins. By contrast, ProCon.org and FactCheck.org aim to indirectly improve quality of discussion by facilitating public access and understanding of trustworthy data [326, 127]. Moreover, those sites do not actually run online debates, as Civil Good proposes to do. FactCheck.org is more politically oriented; its goal is to analyze, clarify, and correct misleading talking points in U.S. political discussion [127]. ProCon.org lists all the pros, cons and facts about given debate topics that are reasonably prevalent in the modern world, aiming for accountability and transparency in the process [326]. Perhaps it is possible for Civil Good to form a symbiotic relationship with these like-minded sites, which can specialize in providing factual information for Civil Good conversations, while allowing the Civil Good website to focus on the actual mechanics of debating.

Other sites have explored the paradigm of debate and serious conversation on the Web, each with its own innovative solutions distinct from those of Civil Good. reddit is a social news and entertainment website that is organized into many small “subreddits.” The Change My View (CMV) subreddit addresses debate on the Web, and has a format where one person states their opinion or belief, and others try to refute it [345]. It is essentially an open devil’s advocate format for debate. reddit’s usual upvotes and downvotes are hidden, replaced by a scoring system called the

delta, which is a record of how many viewpoints a user has successfully been able to change [345]. This is an interesting metric to be considered for use in Civil Good because it is focused on making arguments to change others' views, as opposed to garnering high ratings from others which is an objective that may be prone to system gaming. The CMV subreddit benefits greatly from being a part of reddit, which has a large and diverse community from which the subreddit can draw. However, discussions are not one-on-one and constant moderation is required to keep mean-spirited and derailing comments in check. Since the conversations are held in a public forum, inhibition and group-think are not prevented. Another feature of this subreddit (and reddit itself) is the use of an algorithm that takes timestamps into account to try and bring the most relevant or topical material to the top of the discussion, where it is most visible. This format has been successful enough that the subreddit has amassed almost 100,000 subscribers [345].

Forums for debate, such as DebatePolitics.com and onlinedebate.net, sites typically are well moderated and develop small, tight-knit communities with a culture conducive to debate [90, 301]. Their smaller scale slows the rate of discussion and faces the same difficulties as other group discussion formats like the CMV subreddit. These issues are perhaps even exacerbated, since a smaller group size is more likely to reach a group viewpoint that may prove difficult to overturn [293]. Unlike the subreddit, posts are organized by the time and date they were posted, as opposed to being displayed by a relevance algorithm. This does not seem to be an issue because of the relatively small number of comments to be read. Finally, due to their smaller scale, these forums do not provide significant visibility (at least to the Web at large) for novel ideas and insights that forum users might contribute.

Idebate.org provides another style of online debate with debate-related services like news, debate training, and databases of debates by experienced debaters,

along with an actual debating service [190]. The service is private between two users by default, with the option to allow an audience. The website was designed from scratch to focus on all things debate, instead of trying to adopt any single system for the purposes of debate. Debate.org is yet another solution to the challenge of bringing debate to the Internet [89]. The site offers timed, one-on-one debate with a design that provides debaters with extended time to draft and revise their arguments. Idebate.org and Debate.org do not have a goal of improving public discourse as Civil Good does, but rather are more reminiscent of a structured school debate. However, they are well-designed websites, and they provide a powerful example of keeping users in the right mindset for debate.

Another site, convinceme.net, follows the philosophy that there are many ways to discuss a given topic [77]. convinceme.net offers one-on-one debates that can be viewed and commented on (similar to Idebate.org and Debate.org), as well as group debates (similar to reddit or a forum). More uniquely, convinceme.net offers a competitive “king of the hill” debate format where each user states their position and reasoning on a matter, with the one who best communicates his ideas being declared the winner. Although different from the goals of Civil Good, this approach does provide practice for people to express their ideas eloquently and convincingly. A similar technique might be useful for improving the quality of debate on Civil Good by allowing users to hone their skills.

ForandAgainst.com and CreateDebate.com offer yet additional innovative features to consider [141, 404]. ForandAgainst.com allows for binary yes-or-no voting on various polls constructed by users and for discussion of the results. Unlike ForandAgainst.com, CreateDebate.com has a rough method for keeping score of the conversation and deciding which side of a given two-sided argument is “winning.” The site is also “gamified,” such that users’ scores go up when they contribute to

arguments and when those contributions aid in their side's victory. Although it is more active than ForandAgainst.com, it is still a small site with a relatively tight-knit community. On one hand, this allows for a good debate culture, but on the other hand it does not afford the best possible range of opinions. Notably, CreateDebate.com's relatively simple system has managed to make debate more fun and engaging. The only considerable drawback is that it forces debates to have a "for or against" format, which would not be suitable for topics covering a more gray area. Overall, many solutions have been proposed to solve the problem of getting people on the Web to partake in discourse and debate, with varying degrees of success.

There is another approach for online debate, somewhat different from all discussed so far because it does not actually take the form of a chronological conversation—the approach of iterative development and improvement of ideas by a group. [debatewise.org](http://debatewise.org) illustrates the concept well [91]. It is actually a wiki, a type of website that allows users to collaboratively edit content. [Debatewise.org](http://Debatewise.org) allows its users to document and list all possible arguments for and against topics. This collaborative approach creates one investigative team instead of two competing viewpoints, and the end product is the sum of the group's ideas, rather than the viewpoint of the individual who can best argue their points.

Another site, [TruthMapping.com](http://TruthMapping.com), goes for a more personalized take on this method [411]. [TruthMapping.com](http://TruthMapping.com) allows users to state their beliefs and opinions, map out the premises upon which those beliefs are based, and submit the beliefs in the form of a flowchart to other users of the site. Users then critique the submitted flowchart and the user who submitted it is allowed to revise it until all feedback is incorporated.

Somewhere in between [TruthMapping.com](http://TruthMapping.com) and [debatewise.org](http://debatewise.org) is [DebateGraph.org](http://DebateGraph.org) [312]. [DebateGraph.org](http://DebateGraph.org) allows users to express highly interconnected

topics as a network of ideas, which is revised, elaborated, and improved upon over time, like a wiki. Such a tool has potential for use in brainstorming and mapping ideas or debates, but DebateGraph.org sees comparatively little user activity. That said, DebateGraph.org does include in its user base CNN (the Cable News Network), The Independent, and The White House [312]. Like Wikipedia, these debate-related sites can produce quality research and discussion over time, as opposed to transient debates that are forgotten and perhaps repeated in the future. Accordingly, there is potential for long-term societal learning, assuming such a site could reach a broad audience.

YourMorals.org [313] is not a debate website, but it shares with Civil Good a strategy of appealing to the user’s curiosity. YourMorals.org also shares a similar site flow with Civil Good. It therefore provides a basis for comparisons to find insights that can help improve Civil Good. The home page of the site is simple and free of clutter, and has a concise but informative block of text that manages to give an overview of the purpose of the site, its value to the potential user, and a collaborative appeal to “come join us.” As the first thing users see, the main page of YourMorals.org or of Civil Good should always be something quick and to-the-point that grabs the user’s attention. Civil Good has something similar, a video on the sign-up page, but reading the blurb on YourMorals.org takes only a minute, whereas the video for Civil Good is over twelve minutes long. The site displays a concerted effort towards transparency, as evidenced by the detailed “About” page and the “Links” page to other like-minded sites. This also mirrors Civil Good’s goal of maintaining transparency and trust with the user base. After signing up on YourMorals.org, a user does not receive feedback until they complete surveys and provide data that YourMorals.org can analyze. This too mirrors the way Civil Good tries to provide value to the user only once they have populated their user account

with information. However, Civil Good’s proposed data entry process [252] is longer than the surveys on YourMorals.org. Civil Good would do well to adopt these policies from YourMorals.org. YourMorals.org demonstrates a lot of the planned characteristics of Civil Good put into action successfully, and the differences between the two may help Civil Good learn how to improve.

There is much diversity in the approaches of sites and tools designed for debate on the Web. These sites provide Civil Good with a large data set upon which to draw in terms of what works well for online discourse and what does not.



## 4 Psychology

We examined various facets of the Civil Good website, including the psychology that motivates humans to converse online, how the design of a website subtly shapes its culture, and how the idea of self-affirmation could apply to a website like Civil Good.

We also researched how to create an experience that ensures that users engage with the website on a frequent basis, and how a website like Civil Good could have an immense positive impact on the nature of debate itself.

### 4.1 Online Disinhibition

People behave differently depending on their environmental circumstances [233]. During a religious service, people operate politely out of reverence. During a job interview, people speak and act in their most professional manner to present themselves as strong candidates. Both religious services and job interviews are examples of environments that impose a stricter set of etiquette rules on human behavior than is normal. Environments such as these cause a suppression of behavior known as human inhibition.

Human inhibition plays a role in all facets of living including debate, political or otherwise. Factors like the debate opponent, the audience, and the sensitivity of the subject matter all contribute to the manner in which a person debates. Specifically, people communicate (debate) more effectively when they share similar traits such as age, race, and sex [384, 66].

While debate can often create a relatively inhibitive environment, online discussion has the potential to remove many of the inhibiting qualities of in-person debate. By taking advantage of the nature of online interaction, it is possible to create an environment that facilitates honest and open discussion that allows people

to express their opinions freely and even re-evaluate their beliefs.

#### **4.1.1 Background**

In his paper on the online disinhibition effect, Dr. John Suler brings up six aspects of the Internet that can lower the inhibitions of Internet users [390]. Of these six aspects, Civil Good can leverage two of them: dissociative anonymity and invisibility. Dissociative anonymity relies on the assumption that a person's online identity cannot be readily linked to their offline identity. Even if the two could be linked, there is safety in knowing that they are a single entity in a pool of what could be hundreds of thousands of other users. Dissociative anonymity disinhibits the user by providing reassurance that whatever actions they take will have no repercussions once they go offline. When people no longer believe that they can be judged or punished, it becomes easier for them to speak their opinions, regardless of how controversial these opinions might be.

The second of Dr. Suler's disinhibiting factors is invisibility, which consists of the ability to navigate the Web without both your presence and your activities being known, and of the lack of physical appearance [390]. Human communication is not done solely through words. In face-to-face conversation, there are a multitude of complex social cues that can convey more information than words can. Cues like a sigh, a stutter, or eye contact are all contributors to how people communicate. Whether consciously or not, people take notice of these cues and they affect the direction of conversation.

On the Internet, these subtle cues are often entirely absent. Traditional text communication, the most widely used form of online communication, is one form factor in which the only elements that have a bearing on the conversation are the complete thoughts presented by its participants. Without the hindrances of in-

person communication, debate participants will be able to express their opinions as openly as they like. Non-verbal social cues, like a condescending eye roll or even an approving nod, would be nonexistent. With Civil Good, the only influence is strict and verbal communication.

In addition to the aforementioned cues, readily identifiable information such as race, age, or sex can affect communication in the same way. According to an article by Lee Sproull, people communicate less effectively when they are of different race, age, and sex and vice versa [384]. Sproull also mentions that factors like emotional compatibility between the two parties, the topic of discussion, or the social norms in a particular environment can alter the quality of communication.

Unfortunately, while online disinhibition provides several unique freedoms, it also comes with a corresponding set of unique problems. Online disinhibition comes in two flavors: benign disinhibition and toxic disinhibition [390]. The benefits of online communication discussed so far have relied on benign disinhibition. Under the same influences, people exhibit a tendency to be just as counterproductive as they can be productive. Two more of Dr. Suler's disinhibiting factors have a contribution to this toxic disinhibition: asynchronicity and the minimization of authority.

Online communication has a natural asynchronicity. Nothing is immediate; there is a delay when users type their comments and downtime between each comment. Messages may never reach their intended destination, and there is no way of ensuring that they do. Conversations can end abruptly and at the user's discretion and, because of the lack of immediacy, the consequences of an individual's online actions lose much of their intended impact. Dissociative anonymity has a similar effect in that both dissociative anonymity and asynchronicity contribute to the reduction of consequences on the Internet.

A related contribution to online disinhibition is the lack of attached identities

resulting in a lack of the authority that is attached to those real-life identities [390]. Online, everyone starts off as an equal. A person's offline social standing has no weight once he or she enters an online environment. This creates an environment that supports a greater range of freedom by removing a significant amount of power from authority figures which can have both positive and negative effects on discourse. On one hand, individuals that are not in a position of authority will be able to more easily challenge those who are. On the other hand, toxic individuals gain the opportunity to elevate their perception of themselves increasing their willingness to misbehave. This is the flaw of giving absolutely everyone an equal voice.

In a 1986 field study on electronic communication within a Fortune 500 company, Lee Sproull found that electronic communication caused a sense of status equalization. Comparing messages between employees and their bosses, Sproull found that there were equal levels of uninhibition regardless of whether or not communication went up or down the status hierarchy [384]. Employees were also surveyed on their preferences between electronic communication and face-to-face communication. The results of this survey showed a greater preference for electronic communication when working with their superiors than when working with their subordinates or equals [384].

Toxic disinhibition can be amplified through a process called victim dehumanization. This is the process of mentally removing all human characteristics from a targeted person in order to separate moral standards from the interaction [19] and is most easily recognizable in the form of war propaganda. When the enemy is less than human, it becomes harder to sympathize and easier to do harm. Victim dehumanization is arguably more prevalent in online interaction where there are few reminders of the actual person behind the screen [434].

Toxic disinhibition—indeed, any kind of negative behavior—can start off

a vicious cycle of further toxic interactions, as discussed by the broken windows theory [212]. If “disorderly behavior goes unchecked,” it leads to an increase in further acts that threaten the order of a community—the small acts of disorder build up on one another until the face of the community is utterly transformed, from a “stable neighborhood” to an “inhospitable and frightening jungle” [437]. This follows the idea that people have a tendency to behave in the manner that their environment suggests.

#### **4.1.2 Discussion**

In order to curb the negative effects of online disinhibition, Civil Good must create an environment that makes negative interaction socially unacceptable. The broken windows theory [437] states that people are more likely to engage in negative behavior if negative behavior appears to be the norm. The theory applies just as much in an online environment as it does in an offline environment. For example, when people encounter a forum full of toxic conversation, it is much easier to contribute in a toxic manner since it is the norm in that specific forum [437]. Since there seems to be little to no moderation against that manner of behavior, people can worry less about the consequences. By contrast, when people encounter a forum full of well-mannered individuals, it is difficult to contribute unless they are sure that their contribution will be well-received. This is even more the case when the individuals of that forum have a history of denouncing poor content.

To create such a community, Civil Good could launch with members that already have reputations for producing positive discussion. In Civil Good’s early stages, the developers could invite these members and encouraging them to use the platform and participate in positive discussion [102]. By providing a strong precedent, the community will attract other productive members and drive away

individuals less willing to partake in quality conversation.

Civil Good’s “Seek” concepts may also have an effect in combating inadvertent victim dehumanization. In its original design, a conversation seek requires the user to enter what are potentially extremely specific parameters in order to match the user with their desired conversation partner. By forcing the user to become intimately familiar with their partner’s opinions and social identity, the seek process may act as a powerful reminder of the humanness of their partner. Ideally, this could totally remove the effects victim dehumanization from the Civil Good platform.

Civil Good also reinforces benign disinhibition by refraining from taking any identifying information, including email addresses, from registering Civil Good users. This is because Civil Good’s users must feel entirely confident that none of what they say on the platform will have any effect on their outside lives in order for honest discussion to occur.

#### **4.1.3 Conclusion**

The effects of anonymity and lack of face-to-face contact are often seen as a plague to the users of the Internet. While this is true to some degree, these effects can also be beneficial to services like Civil Good. As long as Civil Good can successfully fight toxic disinhibition, then the benefits of benign disinhibition will come without concern.

## **4.2 Format of Discussions**

For any conversation-based website, it is important to determine the most effective format in which to hold discussions. There are two possible conversation formats to consider: group chat and one-on-one discussion. In order to maximize the efficiency and civility of conversations on the site, it is important to determine which of the

two formats should be Civil Good’s primary means of holding discussion.

#### 4.2.1 Background

In “Small Group Communication: A Theoretical Approach,” [307] Charles Pavitt examines the effectiveness of individual and group decision-making. Pavitt describes three theories on the relationship between group size and productivity. The first is *wholism*: the idea that the collective whole is greater than the sum of its parts, and that performance increases as the size of the group increases. The second theory is *reductionism*: the idea that the whole is only at best equal to the sum of its parts, and that as group size increases, so does the chances of individual members failing to perform optimally, which results in lower overall efficiency. Lastly, the third theory is of the *no-effect relationship*, which essentially states that there is no difference in productivity between a group and an equal number of individuals; that the whole is precisely equal to the sum of its parts [307]. Of the three, Pavitt seems to support the reductionist viewpoint as being the most accurate of the three, arguing that larger groups are less capable of accomplishing tasks together than the individual members of a group are by themselves. In one example, Pavitt cites the work of a researcher named Ringlemann, who performed an experiment to see how much work a group of people put into pulling a rope. Ringlemann’s results showed that as the size of the group pulling the rope increased, the amount of force exerted per person decreased, supporting the reductionist theory, as seen in Table 1 [307].

Similar to the reductionist theory, the theory of *groupthink*, proposed by Irving Janis in 1972, states that rather than come up with their own ideas, members of a large, cohesive group involved in decision-making are more likely to simply accept whatever idea their leader or other group members come up with [413]. One might argue that a larger number of group members provides more variety of information

Table 1: Ringlemann’s Results

Number of Workers	Total Performance	Performance Per Worker
1	1.00	1.00
2	1.86	.93
3	2.55	.85
4	3.08	.77
5	3.50	.70
6	3.78	.63
7	3.92	.56
8	3.92	.49

and ideas, but groupthink theory states that the group will still prefer to reach a quick and easy consensus and will therefore squander that greater variety of information. When a group has adopted a groupthink mentality, its members will often censor their own opinions, selectively process information, fail to adequately explore alternative options, fail to evaluate risks properly, and believe that the group as a whole is good-natured and intelligent enough to not let a bad idea through [413]. Coincidentally, while groupthink theory was both accepted by many following its introduction and presented as a possible rationale for several questionable historical decisions, there is little in the way of empirical research to support it or confirm its effects beyond Janis’ original proposal [413]. Approximately only one study per year has been performed relating to groupthink theory—an abysmally low number compared to the research done on similar concepts in its field. It is therefore very difficult to determine the accuracy of the theory, let alone cite any definitive examples of its presence. In short, the theory holds promise, but the research and support of it is lacking.

In a study performed at the University of California, it was determined that participants in one-on-one chat saw each other more positively in every aspect than those who participated in group chat. In this study, two tests were performed. For the first test, 84 undergraduate students of varying ethnicities were surveyed in order



to determine their degrees of extraversion, agreeableness, conscientiousness, neuroticism and openness. These students were then set up in a number of anonymous, one-on-one conversations with each other and asked to rate their partners using the same criteria, as well as likability, after fifteen minutes of chatting online [256]. For the second test, 72 different students were assigned into groups of six to participate in anonymous online group chat. After fifteen minutes had passed, all members of each group were asked to rate everyone else in the group by the same criteria as those in the first test [256]. The results showed that participants in one-on-one chat saw each other as significantly more extraverted, agreeable, conscientious, stable, open, and likable; as seen in the Table 2 [256].

Table 2: Independent t-Tests for Differences in Perceptions of Participants in One-on-One Group Interactions.

Factor	One-on-One		Group		t
	Mean	SD	Mean	SD	
Extraversion	3.70	0.16	3.25	0.12	8.03
Agreeableness	3.62	0.25	3.06	0.11	7.10
Conscientiousness	3.35	0.17	3.07	0.10	5.05
Neuroticism	2.60	0.13	2.98	0.14	2.19
Openness	3.33	0.16	3.20	0.14	2.19
Likability	4.03	0.34	3.48	0.34	4.10

*Note.* All measures of traits and likability are on a 5-point scale  $df=24$ .  $P < .05$ .  
for t-tests

Another study by a group of University of Chicago students examines the effectiveness of one-on-one tutoring versus conventional group classes. In this experiment, students from grades four, five, and eight with similar pre-experiment scores in certain subjects were randomly assigned to one of three learning environments: conventional group classes, mastery learning (group classes with additional corrective teaching and tests to determine students' mastery of the given subject), and one-on-one tutoring. The experiment ran for three weeks, at which point the final

scores of the students were taken. The results showed that the scores of students who were taught through mastery learning were one standard deviation above those of students in the conventional group classes, whereas the scores of students taught through one-on-one tutoring were two standard deviations higher than those in the conventional group classes [41].

#### **4.2.2 Discussion**

Ringlemann's study supports Pavitt's argument that reductionism is the most accurate theory when it comes to group decision-making. The larger the group was, the less work its individual members put in and the less efficient the group became overall in terms of performing tasks. In order to improve the efficiency of Civil Good, it would be best to limit the size of discussion groups to a small number of people as opposed to allowing large open forums.

If groupthink theory is to be accepted as a genuine phenomenon, then the consequences of it would be that larger groups of people would not in fact provide more diverse information, but would actually be more inclined to conform to whatever the leader or initiator decides. In other words, groupthink would cause large groups to become exceedingly inefficient in terms of development of quality ideas or solutions. For the purposes of Civil Good, allowing an environment where a groupthink mentality can be fostered would be highly toxic to the site. While large groups of people agreeing with each other could potentially be seen as civil, the purpose of Civil Good is to serve as an environment not only where civility can be promoted, but where users can feel free to express unpopular opinions and discuss them openly with each other. If the environment of the site gives rise to a mindset in its users that hinders this kind of open discussion, it would only serve to harm the quality of the site. Therefore, in order to minimize the potential of Civil Good fostering

a groupthink mentality, it is recommended that conversations take place between smaller, temporary groups of users, as opposed to larger, more cohesive groups.

The University of California study suggests that participants in one-on-one discussion view each other far more positively in every aspect when conversing in such a manner. One-on-one discussion is therefore the recommended primary means of conversation for Civil Good since it will maximize the civility of the site as well as the impressions that its users will have of each other.

The data from the University of Chicago study suggests that one-on-one tutoring is by far the most effective means of teaching. One can infer from this that one-on-one tutoring is a far superior means of information conveyance and retention. As a possible parallel, based on this information, one could also argue that one-on-one discussion is a more effective means of conversation than group chat in terms of transferring information to others. It would therefore be in the best interest of Civil Good to focus on one-on-one discussion as it is a more efficient means of discussion.

#### **4.2.3 Conclusion**

These sources all suggest that conversations on Civil Good should be limited to smaller groups of users, preferably one-on-one. Smaller groups of people are arguably superior to larger ones in terms of quality assurance of ideas, effectiveness of information transfer, effort put forth per member, and general attitude towards fellow members. Unless Civil Good limits the maximum number of discussion participants to a small but reasonable number, it would be best for the site to focus on one-on-one discussion as its primary means of conversation so that the efficiency and civility of the site can be maximized.

## 4.3 Reducing Bias with Self-Affirmation

There is a tendency for people to reject information that goes against their personal beliefs and lifestyle choices [369]. People unconsciously discredit this type of information (known as “negative information”) as well as the source of the information [348]. When people hold onto a belief long enough, it becomes one of their defining characteristics. As a result, people see negative information as an attack against their personal image and are less likely to accept that negative information. For example, someone who is a heavy coffee drinker would have a harder time accepting research studies that claim that coffee can increase the risk of cancer. This bias is present regardless of the reputation of the source or the validity of the information [67, 348]. Due to the nature of debate, this human defense has the potential to obstruct effective debate.

The theory of self-affirmation looks to examine the conditions where this bias is reduced. It relies on the idea that people are motivated to protect the integrity of their personal image and states that when a person feels confident about their image, then that person’s bias against negative, threatening information is reduced [370, 369, 285, 79, 348]. In other words, a confident person is more able to objectively analyze information presented to them. In a debate, this is the ideal type of mindset.

### 4.3.1 Background

Many studies examine the effects of self-affirmation by using methods designed to manipulate feelings of self-worth [67, 79, 348, 218, 133]. In one such study, Geoffrey Cohen and associates had study participants examine a report that contradicted their beliefs on capital punishment [67]. Participants that had been given a self-affirming activity beforehand had reported being less critical of the research

compared to participants that were not provided the self-affirming activity. The activities used in this study either asked the participants to write about an important value or gave the participants positive feedback on an important skill. The former activity is known as a values essay [67, 408]. Cohen found that by manipulating a participant's perception of their self-worth, the participant became more receptive to negative information on capital punishment.

Cohen's second study using positive feedback had study participants take a personality test. Participants who were to be self-affirmed had their tests returned to them with 22 out of 25 questions answered correctly and were given a questionnaire labeled "Follow-up Questionnaire for High Scorers." This procedure attempted to boost each participant's positive feelings toward a trait that they identified as important. Participants who were self-affirmed using this method did, indeed, exhibit greater social perceptiveness skills [67]. Cohen's findings provide strong evidence that positive feedback can be an effective source of self-affirmation.

In a study on the self-affirming effects of Facebook use, Catalina Toma and Jeffrey Hancock found that because a user's Facebook profile was dedicated entirely to the positive aspects of that user's life, users had an unconscious tendency to gravitate toward Facebook as a form of self-affirmation [408]. In short, browsing one's Facebook profile, which consists of the user's self-chosen content, will cause increased security in feelings of self-worth.

Steven Fein and Steven J. Spencer published three studies in 1997 examining the relationship between self-affirmation and derogatory stereotyping. They found that participants who had their self-image bolstered were less likely to use negative stereotypes when evaluating other people of a stereotype group. By comparison, participants who were not self-affirmed were more likely to give a negative evaluation. In addition to this, participants who had their self-image threatened reported

having an increase in self-esteem suggesting that negative social comparisons with stereotype groups are a common form of self-affirmation [133].

On the subject of awareness of self-affirmation, Sherman et al. published three studies looking at whether or not self-affirmation could occur unconsciously, whether or not self-affirmation effects were diminished when participants were aware of the effects, and whether or not the effects were diminished when participants were aware of a connection between self-affirming activities and the receptiveness of threatening information. In short, they determined that self-affirmation was an unconscious process whose effects were reduced when there was awareness of its effects and when there was awareness of its connection with threatening information [371].

#### **4.3.2 Discussion**

Based off of existing self-affirmation research, here is a set of guidelines containing five criteria for developing an effective self-affirmation manipulation method:

- The method should make use of a domain that is important to the user. For example, a values essay about one's skill in American football will be ineffective to someone who is not a football player [408, 369, 79].
- The method should not promote awareness of its effects to the user. Self-affirmation should be an unconscious process [408, 370, 371]. If the user is aware that its purpose is to reduce bias, its effects could be diminished [371].
- The method should be unrelated to the negative information that is later received [369, 285].
- The method should be general and should not require tailoring for each individual person. This is especially true for methods that attempt to take advantage of mass communication [408, 285].

- The method should, ideally, offer a control equivalent in order to measure its effectiveness [285].

These guidelines should help Civil Good leverage the benefits of self-affirmation in order to develop a strong, receptive community.

One of Civil Good’s original example implementations of self-affirmation theory is to have users write a values essay before participating in conversation [252]. The values essay is a manipulation method that has been repeated many times in multiple studies on self-affirmation [67, 370]. It is a proven method for causing the desired effects of self-affirmation and also meets the first criterion listed above. By allowing the user to choose the essay topic, all guesswork on user-relevance is eliminated. Whatever the topic is, it will be relevant to that specific user.

Unfortunately, a values essay implemented in this way also has several disadvantages and, in the original example, violates the criterion stating that self-affirmation should occur unconsciously. By putting the exercise directly before initiating conversation, an uninformed user may question the purpose of such a seemingly unrelated exercise. To the average user, the exercise only impedes the Civil Good experience as a whole. By contrast, an informed user might not receive its intended effects if they are aware that the exercise is there to reduce personal bias [371]. In either case, having a single static exercise before every conversation will eventually lead the user to comprehend its intent or cause the exercise itself to become a routine that serves no purpose other than to waste time.

These problems are only compounded when considering the fact that the essay exercise takes place after the already lengthy conversation search process (known as a “Seek” in Civil Good.) It adds another hurdle between the user and the desired debate and does not have an explicit contribution to the end goal. On top of that, a user approaching debate time is likely focused on the debate topic. The timing of

this exercise could make it an unnecessary distraction.

Attempts have been made to address this issue by reducing the obviousness of the affirmation. Civil Good could take a different approach and use positive feedback as a self-affirming mechanism. Something as simple as a congratulatory notification might be enough. If this were the case, there would be no need to have a forced exercise. For example, Civil Good could periodically notify User A that User B has changed their opinion to align with User A's original argument. This would give User A a sense of importance in the Civil Good community; it would reinforce the idea that the user's voice has been heard and that they are making a difference. Another more obvious example might be to congratulate users on improved ratings. Both of these types of notifications utilize positive feedback, supported by Cohen's second study discussed in Section 4.3.1, as a source of self-affirmation. Giving the user a sense of progression, using this form of positive feedback or otherwise, could be a good way to improve their experience on Civil Good and may even assist in user retention.

However, these examples come with their own set of problems. The first criterion states that self-affirmation relies on bolstering an important aspect of the user's life. If a user has little interest in Civil Good, there might be little reason to praise their Civil Good performance. Another issue is that because reading a congratulatory message is entirely voluntary, the user could potentially never even see the positive feedback message intended to improve their behavior.

To handle these issues, Civil Good could consider making self-affirmation activities entirely optional, but encourage their use through the natural benefit of these activities. If Civil Good were to attempt to mimic Facebook's effects discussed in Section 4.3.1 by keeping a "Civil Good profile," self-reflective users might be self-inclined to browse their profile and, thus, reaffirm themselves. This could lead to a



user base that is more likely to be in a receptive state of mind. For example, Civil Good could keep a history of a user’s conversations and, by default, display the most well-received conversations in the forefront. Civil Good could also display the user’s achievements, like conversation milestones, in the same fashion. Hopefully, the user is proud of these conversations and achievements and considers them part of their personal success.

Another possible solution could be the management of an entirely separate conversation system. Remove debate from the equation and allow people to discuss their hobbies, interests, skills, or experiences. This takes the original idea of a values essay and puts control into the hands of the user. According to a 2007 quasi-experiment on forced-response in online surveys, there is an approximate 49.1% dropout rate when users are forced to answer an online survey [387]. Compared to a 21.2% dropout rate for voluntary surveys, this kind of voluntary values essay should have a greater completion rate than the previously forced values essay.

### **4.3.3 Conclusion**

Self-affirmation manipulation could be a powerful tool in fostering a positive community. It is proven to improve receptiveness within stubborn individuals. If used correctly, it could act as a deterrent against toxic discussion on Civil Good and enhance the service’s overall quality. If ignored, discussion quality could degrade with people’s natural tendency to utilize negative stereotyping as an ego boost [133]. This is even more of a threat for a service like Civil Good which relies on stereotyping (through search parameters) to identify conversation partners. Following the guidelines in Section 4.3.2, Civil Good has an opportunity to take advantage of this psychological concept in an environment where it would be most beneficial.

## 4.4 Other Psychological Influences

In Section 4.3, self-affirmation was discussed as having a reduced or insignificant success when the user was aware of its psychological effects. Other psychological principles attempting to explain bias and persuasion might also be affected by awareness of the principle. Based on awareness' effect on self-affirmation, awareness of negative influence might be beneficial to logical thinking.

### 4.4.1 Background

Fritz Strack et al. published an experiment in 1993 examining the effects of awareness of basic influences on judgment [389]. The study had participants go through three seemingly independent studies, the first of which was a listening exercise in which participants listened to a recording of a set of words and wrote down what they heard. Some participants were given a recording with positive words while some were given a recording with negative words. All participants were subsequently given an exercise in which they were to judge the ambiguous actions of a fictional character. Some participants were reminded of their earlier listening task. Those who were not reminded of the prior task exhibited greater assimilation with the unconscious positive/negative influences in their judgment than those who were. However, some participants who *were* reminded seemed to overcompensate when re-evaluating their judgment by intentionally ignoring the irrelevant information from the listening task. These results suggest that awareness of irrelevant influences does reduce the effect of that influence [389].

The boomerang effect is a psychological phenomenon in which people react to information in a way that is opposite the intention of the information provider [440]. The psychological concept of reactance explains this behavior as a person's attempt to exercise their attitudinal freedom when they are presented with information that

potentially restricts their freedom to think for themselves [440, 438]. In 1998, David MacKinnon and Angela Lapin attempted to test the boomerang effect by looking at the reactions of alcohol users and non-users to alcohol advertisements and health warnings [249]. While they largely found no evidence of the boomerang effect, MacKinnon and Lapin did discuss one other possible cause of the boomerang effect: cognitive dissonance [249]. When presented with information on the negative effects of alcohol use, drinkers may feel a sense of discomfort and may unconsciously attribute more benefits to alcohol use in an attempt to curb this discomfort.

Another study looking at the effects of anti-smoking messages was conducted by Joyce Wolburg in 2006 [438]. Wolburg found some evidence of the boomerang effect amongst smokers exposed to anti-smoking messages. However, some smokers in the same category exhibited an increased intention to quit. For non-smokers, their decisions were only reinforced. This suggests that the boomerang effect is influenced by the strength and direction of the individual's predisposition to the advocated attitude.

The halo effect is another source of judgment bias in which people's global evaluation is affected by their evaluation of smaller independent parts [292]. Richard Nisbett and Timothy Wilson's conducted an experiment in 1977 in which participants were presented with one of two videotaped interviews: one where the interviewee behaved in a friendly manner or one where the same interviewee was instead cold and distant [292]. All participants were asked to rate the interviewee on appearance, mannerisms, and vocal accent. Nisbett and Wilson found that participants who were given the friendly interview rated the interviewee more positively than those who were given the cold and distant interview. These results provide ample support for the halo effect theory. Nisbett and Wilson also made sure to note that all study participants were unaware of the halo effect.

#### 4.4.2 Discussion

As discussed in Section 4.4.1, Strack et al. determined that increased awareness of irrelevant influences on judgment will result in re-evaluation of the judgment. While Strack only examined a single influence (recordings of positive or negative words), such irrelevant influences can also include unintentional biases coming from the boomerang effect or the halo effect. By extension, the boomerang effect and halo effect could be reduced through education and increased awareness of the resulting bias. This is consistent with the effects of awareness on self-affirmation discussed in Section 4.3. People who are more aware of the bias reducing properties of self-affirmation receive a diminished influence from self-affirmation, just as people who are more aware of the boomerang effect or the halo effect may receive diminished influence on their judgments from those phenomena.

Increased awareness of bias-inducing concepts like the boomerang or halo effects will be beneficial for Civil Good, as awareness has been shown to reduce the effects of these biases. At the same time, Civil Good should also be keep in mind the potential repercussions of increased awareness. Strack et al. warn that people who are aware of these influences may overcompensate when recomputing their initial judgments. When people re-evaluate, they will intentionally disregard the irrelevant information (such as the biases caused by the boomerang effect) and may produce a judgment that is more extreme than if they had not been exposed to that information (or bias) in the first place.

Should Civil Good choose to attempt to mitigate these adverse effects, it is recommended that the service only use light reminders of the potential biases. In Strack’s experiment on influence awareness, participants in the “reminding” condition (those who were reminded of their initial exercise) only required a 40 second survey of their previous listening exercise to cause them to adjust their judgments

accordingly. As long as Civil Good users are initially knowledgeable about their possible sources of bias, they should only require a short mention of the sources in order to bring appropriate attention to their own biases.

#### **4.4.3 Conclusion**

Civil Good would benefit from increased awareness of the causes of negative biases. With reduced influence from these biases, people should be able to more logically determine their personal opinions and more effectively evaluate the arguments of their peers. However, caution is advised with this approach since the possibility of overcompensation exists amongst individuals aware of the source of their biases.

## 5 Legal Issues

There are legal issues arising from the publication of content on a social media website, such as privacy issues, intellectual property rights, and litigation avoidance [336]. This section discusses the major legal issues faced by United States-based social media websites: the handling of personally identifiable information, the handling of content protected by intellectual property rights, defamation issues, and the two major United States legislative enactments governing Internet content, the Digital Millennium Copyright Act and the Communications Decency Act. Civil Good needs to determine the extent to which it must address international legal restrictions and requirements, but international laws are outside the scope of this section. Due to the relative recency of the Internet and social media and the rapid evolution of communications and information storage technologies, the principles governing these legal areas are still young and evolving, and require continual monitoring to keep up with new developments. The current status of these areas is summarized below.

### 5.1 Personally Identifiable Information

Perhaps the most prominent privacy issue facing social media sites is the handling of users' personal information—what information to collect, how to protect it, and what steps to take in the event of unintended disclosure. Such information has come to be identified by the phrase “personally identifiable information” (“PII”).

#### 5.1.1 Background

PII is defined by the National Institute of Standards and Technology as “information about an individual including (1) any information that can be used to distinguish or

trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information" [259].

Other definitions exist, such as the broader provision contained in California's recently enacted "Do Not Track" law: "individually identifiable information about an individual consumer collected online by the operator from that individual and maintained by the operator in an accessible form, including any of the following: (1) A first and last name; (2) A home or other physical address, including street name and name of a city or town; (3) An e-mail address; (4) A telephone number; (5) A social security number; or (6) Any other identifier that permits the physical or online contacting of a specific individual" [279].

State laws that govern the protection of PII include the California Online Privacy Protection Act of 2003 [244], Nevada Revised Statutes 603A-Security of Personal Information [231], and 201 Code of Massachusetts Regulations 17.00: Standards for The Protection of Personal Information of Residents of the Commonwealth [298]. Consumers have an expectation that website operators will take appropriate security measures to ensure the protection of their PII.

At least 46 states have enacted statutes requiring notification of a security breach involving PII [386]. These laws have varying requirements as to what constitutes PII, what constitutes a security breach, and what notification actions have to be taken in the event of a breach.

In addition to ensuring non-disclosure of PII, websites must also be aware of the need to properly respond to user Web browser "do not track" mechanisms, which seek to prevent websites from using cookies and other constructs to monitor user Web behavior [143]. California recently enacted a law requiring websites to

disclose how they respond to browser “do not track” signals [279].

The U.S. Federal Trade Commission has published “The Fair Information Practice Principles” (FIP), a set of recommendations for businesses’ use of PII. The FIP recommends that websites give notice of their PII policies prior to collecting PII, that users be permitted to decide how their PII is used, that users be able to access their PII to determine its accuracy, and that websites take necessary steps to protect the security of PII [69].

### **5.1.2 Discussion**

To accommodate differing definitions in federal and state laws regarding what constitutes PII, Civil Good should take a broad view as to what may be considered to be PII. The current draft of Civil Good’s privacy policy provides that Civil Good will not collect any PII other than IP addresses and email addresses, and will treat that information as PII. Appropriate safeguards (such as encryption) are required to protect that information.

While users will be providing personal information such as religion, political views, education history, and employment, that information may not actually constitute PII. In order for information to be considered PII, Civil Good would need to have the ability to connect that information to the user’s identity. Given that email addresses will not be linked to accounts, this may not be possible. However, if it is possible to connect the information to an identifiable person due to the storage of IP addresses or other data, then this information may be considered PII. Accordingly, Civil Good should treat all such information as PII and protect it as such.

Civil Good will be required to continually develop and publicly post a privacy policy clearly specifying what PII will be collected from users, how collected PII will be utilized by Civil Good, how PII will be protected against disclosure, and what



Civil Good will do to notify users in the event of a PII security breach. Because Civil Good at present does not plan to link email addresses to accounts, there may be logistical difficulties in complying with state breach notification laws, and Civil Good would require legal counsel to review these statutory requirements and determine the policies Civil Good would have to implement to ensure compliance. Once a notification policy has been determined, it should be posted on the website as part of the PII policy notices. Civil Good will also be required to develop, implement, and publicly post a policy explaining how the site will deal with browser “do not track” signals.

The states of Alabama, Kentucky, New Mexico and South Dakota have no laws specifically pertaining to security breach notification [68]. Apart from these, all states require some form of written notice to any individual whose PII may have been compromised in the event of a security breach, and most states require that the disclosure be made “in the most expedient time possible and without reasonable delay” [68]. Some states, such as Arkansas and Colorado do not require a disclosure notice of privacy breach if a good-faith and reasonable investigation determines that user PII has not been compromised [68]; however, it is recommended that Civil Good notify all users of a potential privacy breach regardless of the results of the investigation through the best method of communication available (via email when possible; otherwise, via a post on the homepage or team blog). Civil Good should still seek legal counsel to determine whether its policies regarding the notification of security breaches is compliant with all state laws.

It is also recommended that Civil Good comply with the Federal Trade Commission’s “Fair Information Practice Principles” and post its compliance with these guidelines.

### **5.1.3 Conclusion**

A website's failure to protect PII is a breach of trust, and is treated by the media as such in the all too frequent news stories recounting the latest disclosure. Many users will no doubt express intensely personal feelings on Civil Good, and improper disclosure of their PII may be deeply resented by such users. For these reasons, Civil Good needs to treat PII collection and protection issues as crucial to its successful growth.

## **5.2 Intellectual Property**

An important issue facing social media websites that permit the public posting of user-generated content is whether the publication of the content violates others' intellectual property rights, usually copyrighted material.

### **5.2.1 Background**

Federal copyright law protects "original works of authorship" and gives the author a seventy-year exclusive right to all publication of the work. Violations of another's copyright can lead to liability for monetary damages as well as injunctive relief (an order from a court to take some affirmative action or not to act in a certain manner) [334].

However, Section 107 of the Copyright Act protects the "fair use" of copyrighted material, such that a social media website may not be liable for publication of copyrighted material if the use is "beneficial to society" and if the site is not profiting from the use [328].

The Digital Millennium Copyright Act (DMCA) is a Federal law enacted in 1998 governing the publication of creative works on the Internet. Section 512 of

the DMCA establishes a “safe harbor” provision, whereby Internet companies can shield themselves from copyright claims that could arise from third parties posting copyrighted works on the website [338]. In order to take advantage of these safe harbor protections the website must implement and follow a “notice and takedown” procedure, by which copyright owners can notify the website of alleged copyright infringement and have the offending material removed [331].

Section 512(c) of the DMCA states that the operators of websites are not liable for monetary damages from an alleged copyright violation so long as they did not know (or should not reasonably have known) of the violation, did not financially benefit from the violation, and quickly removed the violating material upon receiving notice of its presence [73]. Section 512(d) contains similar protections for instances where the website includes links to other sites containing infringing copyrighted material [73].

A trademark is a word, phrase or symbol that uniquely identifies a business or organization, such as “Google” or “Microsoft.” Trademark owners have the exclusive right to use their trademarks, and can prevent others from using them in a manner that may confuse consumers [333]. Internet content publishers must be aware of potential liability for third parties generating content that violates another’s trademark rights.

### **5.2.2 Discussion**

Because some conversations within Civil Good will be publicly searchable, Civil Good will be required to take steps to avoid liability for third parties posting material that may infringe upon another’s copyright. This is governed by the DMCA, and in order to comply with the DMCA, Civil Good will be required to develop, implement and post a “notice and takedown” procedure for copyright owners to make claims to

Civil Good of alleged copyright violations and to respond to those violations. Legal counsel may be required to determine whether a copyright claim is valid, and all infringing material should be immediately removed from public view.

In addition, Civil Good should remove all reported infringing material, even if a violation notice is not received. If there is any financial benefit to Civil Good from the posting of publicly available material, Civil Good should review the material before posting to ensure it does not contain copyrighted material. Civil Good should also post a link to the DMCA and publicly post Civil Good's policy for compliance with it.

The key issue with regard to trademark infringement is whether the publication of the trademark is likely to lead to confusion of the consumer [332]. Due to the nature of Civil Good's conversations, there does not seem to be a strong likelihood that conversations will contain material infringing on third parties' trademark rights, but Civil Good should be aware of the principles governing the issue and be responsive to complaints of infringement.

### **5.2.3 Conclusion**

Because Civil Good does not itself generate content, it can avoid liability for the posting by third parties of others' intellectual property by following the safe harbor provisions of DMCA. This will require the posting and implementation of a compliant "notice and takedown" procedure. So long as Civil Good promptly complies with "notice and takedown" requests, it should not face any liability for such content.

## 5.3 Defamation

Debates on issues always bring with them the possibility of personal attacks on the debater or on public figures. The occurrence of such personal attacks raises the issue of whether Civil Good might face liability for defamation.

### 5.3.1 Background

“Social media websites must guard against the possibility that user generated content will be defamatory towards another person. State laws vary on what exactly constitutes defamation” [327]. Generally speaking, defamation occurs when a person publishes a false statement about another that harms the reputation of the subject of the statement, and the person making the statement did not have any special legal privilege to do so [327]. If the subject of the statement is a public figure, defamation usually requires that the false statement have been made with knowledge of the falsity of the statement or reckless disregard for its truth or falsity [330]. Only statements of fact can be defamatory, statements of opinion cannot be defamatory [329].

Section 230 of the Communications Decency Act shields websites from liability for certain types of claims based upon the website publishing content created by third parties [337]. The immunity granted by Section 230 extends to defamation and invasion of privacy claims, but does not cover intellectual property claims. Under Section 230, website operators are permitted to passively host content created by third parties, exercise “traditional editorial functions” over such content, screen and remove content, and encourage third parties to submit content. However, if the website operator changes the meaning of the content, it may lose the immunity [335].

### **5.3.2 Discussion**

Liability for defamation does not require broad publication. Publication to only one person is sufficient; however, broader publication would likely increase potential damages [327]. Due to this, liability for defamation could arise from a one-on-one conversation as well as from the content contained in publicly searchable conversations.

To protect against liability for defamation, Civil Good should comply with Section 230 of the Communications Decency Act, which permits passive hosting of third party-created content that may be defamatory but prohibits active editorializing of the content that significantly changes the meaning of the content. Civil Good may engage in what is considered to be traditional editorial functions such as deciding whether to publish or remove content and make minor editorial changes (for example, removing profane language) [335].

### **5.3.3 Conclusion**

Civil Good can protect itself against what is likely a small danger of liability for defamation by complying with Section 230 of the Communications Decency Act, which requires that Civil Good not meaningfully alter the content of potentially defamatory material posted by third parties.

## **5.4 Information Requests**

Websites that collect PII and publish political views of its users must be aware of potential interest in the identity of the users who post opinions potentially impacting criminal law. For example, a user posting a perceived threat towards a political official may draw governmental attention. Civil Good will be required to properly

address and handle information requests that may arise from such postings.

#### 5.4.1 Background

Recent news reports have publicized numerous instances of governmental information gathering, including PII, from various types of companies, including website operators [167]. In many instances, the collection of PII was pursuant to formal legal means, such as a subpoena or court order [32]. A subpoena is a document commanding the recipient to produce information, which in many cases can simply be completed and served by the issuing agency without any court approval [136].

Under numerous court decisions and governing statutes, including the Electronic Communications Privacy Act (ECPA), the PATRIOT Act, and the Foreign Intelligence Surveillance Act (FISA), governmental agencies can compel a website operator to provide various types of electronically stored user information without first obtaining a warrant or court order. For example, under the ECPA, a federal agency can obtain emails older than 180 days by simply sending a subpoena to the subject website [270].

Court decisions have been liberal in permitting governmental access to information. For example, the United States Supreme Court in 1979 ruled that installation of a “pen register” that recorded incoming and outgoing telephone numbers was not a “search” under the Fourth Amendment, and thus did not require a warrant [1]. In 2007, the United States Ninth Circuit Court of Appeals, applying *Smith v Maryland*, ruled that tracking of IP addresses also is not a search under the Fourth Amendment and therefore can be done without a warrant [2].

It appears that government agencies have been able to obtain extremely broad orders for collection of numerous types of information under FISA [10]. Traditional court orders can be challenged, and it appears that even FISA court orders

may be subject to challenge [158].

There is strong evidence that website operators such as Google, Facebook, Apple, and Yahoo have voluntarily provided information in response to governmental information requests made without legal process [248, 325]. There is a vigorous ongoing debate as to the extent to which the law authorizes this formal and informal collection and use of PII by the government [166, 33].

There have also been revelations that the National Security Agency (NSA) intercepted unencrypted customer data from Google and Yahoo by tapping into those companies' fiber data lines connecting their data centers, without notice to Google or Yahoo [147].

#### **5.4.2 Discussion**

Because Civil Good does not link a user's email address to their accounts, it should be difficult to connect a user's supplied PII with their identity. However, it will not be impossible since IP addresses, unique usernames, or other information may be available to link PII to particular individuals. Accordingly, governmental authorities may believe it useful to seek PII from Civil Good.

Civil Good will require procedures to deal with such requests. Legal counsel knowledgeable in the field will be required to review subpoenas and Court orders to determine their validity, whether to challenge them in Court, and what, if any, information to provide in response. The review will have to evaluate the type of information sought, the claimed justification for the request, the user privacy interests impacted, and the cost of challenging the request. Civil Good will be required to comply with the subpoena or court order to the extent that it is deemed valid. Civil Good's privacy policy should clearly state that Civil Good will provide information in its possession when legally required.



Informal information requests will have to be evaluated and handled as a business decision. If there is no legal compulsion to produce information, Civil Good will have to determine whether it should in all cases decline the request in order to preserve its users' privacy or whether the particular circumstances of a request warrant release of information. This is an ethical and business decision, not a legal issue, which Civil Good's operators would have to determine.

The issue of governmental information requests for electronically stored customer data is complex and evolving. Because such information requests can be so varied and fact-specific, such requests will require individual handling as they are received.

### **5.4.3 Conclusion**

The likelihood of Civil Good receiving a governmental information request is unknown. However, given the probability of users posting views of interest to the government, Civil Good at its inception should have a relationship with legal counsel with knowledge and experience in such matters, so that all such requests are properly handled.

## **5.5 Use by Minors**

Because Civil Good will no doubt host conversations on topics not suitable for minors (persons under the age of 18), Civil Good will need to be aware of legal issues addressing the exposure of such material to minors and will be required to formulate procedures governing minors' use of the site.

### **5.5.1 Background**

The Children’s Online Privacy Protection Act of 1998 (COPPA) is a federal law regulating websites’ collection of PII from children under 13. The act governs parental consent requirements, what information can be acquired, security requirements, and privacy notice requirements [299].

### **5.5.2 Discussion**

In order to limit the exposure of inappropriate material to minors, we recommend requiring users to specify their age by selecting from a list of age ranges and limiting the content that minors are allowed to view. The ranges should be chosen such that children under 13, minors between 13 and 18, and persons over 18 are in different age ranges (e.g. <13, 13-17 and 18-25). Children under the age of 13 should be prohibited from using Civil Good so that Civil Good will not be required to comply with the many complex requirements of COPPA.

In order to reduce the chances of being sued, users who report to be under 18 should be prevented from searching for, reading, and participating in any conversations pertaining to alcohol, firearms, tobacco, drugs, sex, and related mature topics. In addition, each conversation page should have a report button to flag it as for mature audiences only. It would also be sufficient to require users to declare that they are 18 or older in order to participate in Civil Good at all [39]; however, this may result in a loss of potential users. Regardless, legal counsel should be retained to ensure there are no state laws requiring additional efforts.

### **5.5.3 Conclusion**

Civil Good will need to take measures to avoid its improper use by minors and to attempt to prevent its use by children under 13, as any involvement with inap-

propriate contact with minors could be harmful to the continued operation of the site.

## **5.6 General Litigation Avoidance and Defense**

There may be other state laws governing the operations of social media websites that would impact Civil Good. For example, proposed North Carolina and Connecticut bills would require social networking websites to verify users' ages and obtain parental consent for minors [131]. Civil Good would need legal counsel on retainer to advise of such laws and to keep abreast of newly enacted laws requiring sites such as Civil Good to adopt certain procedures.

In addition, Civil Good should clearly designate an individual who is authorized to act and speak on behalf of Civil Good and implement policies preventing other workers from claiming such authority so that all actions taken on behalf of Civil Good are only taken by those in position to perform them correctly [353].

## 6 Societal Impact

We researched the societal impacts a website like Civil Good could have and what mass behavioral changes it could foster.

### 6.1 Political Polarization

Civil Good will be looking to make use of research underlying the causes of political polarization in order to improve debate. This topic of political polarization in America has seen a growing trend in recent years. While there has certainly been an increase in polarization between members in a position of power (known as “elite polarization”), there has been much debate over whether or not polarization has actually increased amongst the masses (known as “mass polarization”) or whether it is just a perceived increase due to growing media attention [228].

#### 6.1.1 Background

Marc Hetherington’s 2001 political science review found evidence suggesting an increase in mass polarization using a “feeling thermometer” survey (a method that typically has respondents apply a hard numeric value between 0-100 based on their personal feelings [286]) that noted a 40% increase in partisanship among respondents [180]. According to his findings, he argued that mass opinion is shaped in response to changes in elite opinion. In other words, if the elite become more politically polarized, the masses will soon follow.

Hetherington also found that there has been a steady rise in elite polarization starting with the 95th Congress in 1977 [180]. This was done by calculating each House member’s DW-NOMINATE scores (a score meant to evaluate each member’s relative political standing developed by Keith Poole and Howard Rosenthal [321])

and the mean Euclidean distance between each opposing party's scores [34]. For the 95th Congress, Hetherington calculated a distance of 0.448 [180]. By the 104th Congress, this rose to 0.748. Using these findings, Hetherington determined that politics at the elite level experienced increased polarization during a period slightly preceding increased polarization among the masses. This evidence supports his claim that elite politics act as a guide for mass politics.

To explain opposing studies that suggest that mass opinion remains neutral irrespective of changes in elite political polarization, Alan Abramowitz claims that only politically-interested individuals will react to changes in elite politics [8]. He argues that much of the American public is not politically interested and, thus, will not react to elite polarization.

Looking instead at the causes of individual political extremism, Philip Fernbach conjectured that politically extreme views tend to have their roots in the overconfidence in understanding [134]. According to Fernbach, people take more extreme views of political policies when they believe they understand the policy better than they actually do.

Fernbach's experimental procedure had study participants provide a mechanical explanation of the selected political policy [134]. He argued that by having people actually attempt to describe these policies, it would force them to confront their lack of understanding and, as a result, take more moderate stances on the issue. By forcing people to come to realize their ignorance on the subject, they become less confident in their original standing as it was made while they were significantly under-informed.

This is supported by Patricia Linville's 1982 study on the relation between depth of examination and extremity of evaluation [241]. Linville presented each participant with five different cookies. She then had participants perform one of

two procedures: (1) evaluate two attributes of their given cookies or (2) evaluate six attributes of their given cookies. Each participant then marked their overall impression of the cookies on an unmarked 5-inch scale. Linville found that those who had examined fewer attributes gave more extreme overall ratings (greater standard deviation) than those who had examined all six attributes. Linville’s findings create a parallel to Fernbach’s claim that those who try to examine political policies in more detail will have less extreme opinions.

### **6.1.2 Discussion**

In order to reduce negative discourse caused by party polarization, polarization derived from ignorance must be mitigated to the fullest extent possible. Since, according to Fernbach’s study in Section 6.1.1, most politically extreme views have their roots in ignorance and overconfidence, the obvious solution is to combat this ignorance. With an informed community, extreme political views will be minimal and will allow for more productive debate about personal preferences and positions.

Fortunately, Civil Good already has a mechanism planned for informing its community. The original design document [252] suggests a conversation type called a “Query” where “the Seeker asks for enlightenment, instruction, or explanation from another member.” While it is not the ideal way of getting informed, it does fall in line with Civil Good’s original intent. Civil Good should encourage users to seek up-to-date information, either through “Query” conversations or some other means, before getting into an actual debate. This could potentially improve overall knowledge on the topic and push people to do learn more about their debate topic.

### **6.1.3 Conclusion**

Political polarization is an occurrence that often threatens to damage constructive discourse in real-world debate. Being an online tool, Civil Good has the chance to guide its users toward less extreme mindsets through education. Both a more neutral and more informed community can only benefit Civil Good.

## **6.2 Minority Opinion Representation**

The “spiral of silence” theory posits that popular opinion drowns out minority opinion, creating the illusion that popular opinion is more prevalent than it actually is [293]. Specifically, people who hold a minority opinion tend to be more reserved about expressing that opinion in order to continue being socially relevant [293]. As a result of their silence, the majority opinion seems more substantial with vocal supporters unconcerned about becoming socially isolated. Since popular opinion appears to have greater support, other individuals who hold the same minority opinion will also remain silent, hence the term “spiral of silence.”

With the current state of discussion online and offline, those with a minority opinion are given few options to voice that opinion without putting themselves at risk of social rejection or reprisal. In almost all cases, there is some way to link the user’s real identity with their arguments online. Civil Good is an attempt to effectively provide a portal in which those in the minority can be given equal representation. The minority opinion needs an equal voice for fair and insightful discussion.

### **6.2.1 Background**

As social beings, humans have a fundamental need to be a part of a group and to have some form of interpersonal relationships [36]. People within a group are

faced with pressure to conform with the group. In a 1990 experiment on group conformity, Dominic Abrams et al. replicated Solomon Asch's original conformity experiments in which participants were presented with the image of a line and were asked to determine which of three other lines were of equal length to the original image [9, 179]. Each participant was put through 18 trials (18 different lines) and grouped with three confederates whom the participant believed were also study participants [9]. The confederates were trained to provide public, unanimous answers that were either correct or erroneous. The actual study participant always answered last after hearing the answers of the three confederates.

Abrams and associates recorded a rate of 77% conformity with an erroneous answer among all participants [9]. That is, 77% of all participants agreed with an erroneous answer on at least one trial. Of 432 total trials where the confederates gave an incorrect answer, participants conformed during 138 trials.

Another replication of Asch's experiments by Deutsch and Gerard tested the hypothesis that people are less likely to conform when they do not perceive pressure from a group to conform [97]. To test this, Deutsch and Gerard compared conformity rates between participants who were face-to-face with confederates and participants who were in private away from confederates. They determined that people are more likely to conform in a face-to-face situation than in a private situation [97].

Solomon Asch's original series of experiments also examined the effect of the size of the majority group [28]. Asch varied the number of confederates between one and fifteen. He found that there was a slight influence when participants were opposed by a single person. This influence became more significant when there were two confederates present and even more significant with three confederates. Beyond three, there was no notable increase. Asch concluded that the size of the opposing majority group did have an effect, but only up to a certain point [28].



Looking at group conformity on more socially and personally significant attitudes, Matthew Hornsey et al. published a paper in 2003 examining different levels of group conformity. In this domain, they found evidence supporting the observed assimilation of minority attitudes to the majority [183]. Using information gathered through a survey of students who held opinions resisting the political state of affairs at the time, they concluded that people with a weak basis for their attitude are likely to conform with the group norm either because of the uncertainty in their attitude or because of the benefits gained from group acceptance. It should be noted that this study only examines the intent of individuals and not actual action within a group setting. While intent does not necessarily translate into action, the fact that intent is affected supports the overall argument that the majority can wield over the minority.

In 1987, Martin Kaplan and Charles Miller conducted an experiment examining the dominance of either normative influence (influence that leads people to conform in order to comply with the majority) or informational influence (influence based off of facts) in group discussion [210]. In groups of six, they had participants discuss a trial of civil damage where they were tasked with determining the plaintiff's award for either compensatory damages (direct payment for the damage of property) or exemplary damages (payment extracted from the defendant as punishment). Kaplan and Miller hypothesized that the discussion over exemplary damages, being a situation where there was no exact calculable value for the award, was a case where normative influence would predominate in the persuasion of group members. Their experiment determined that in discussions where the focus is on judgmental issues, normative influence are, indeed, the predominant influence.

The relative popularity of one's opinion (known as "opinion climate") is a primary factor toward how willing one is to express that opinion [258]. Jörg Matthes

et al. published three studies in 2010 examining the relationship between the level of conviction with which people hold an opinion and their willingness to express that opinion. They found that people who were not absolutely certain of their opinion were, indeed, affected by opinion climate while those who were certain were unaffected [258].

### **6.2.2 Discussion**

To accommodate for those with unpopular political attitudes, Civil Good aims to reduce group pressure using one-on-one debate as opposed to public group debate. With significant research supporting human tendency to either conform with the majority or remain silent when facing heavy opposition [258, 293, 183, 9, 179], the political minority currently has bleak prospects for accurate representation. In most present situations, only the most adamant of the political minority has strong enough conviction to speak against the majority [258].

One-on-one debate also accommodates those who are uncertain of their opinions and who often look to the majority to guide their uncertainty. Within the scope of a single debate, a majority opinion only exists if there is complete agreement between both parties. There is no majority to look to for guidance in this type of environment. However, there is also no threat of social exile, leaving both parties are free to explore each other's viewpoints without fear of judgment.

Additional evidence supporting Civil Good's style of discussion comes from Kaplan and Miller's experiment discussed in Section 6.2.1 where they concluded that debate over judgmental issues are primarily steered by normative influences. Such judgmental issues include the proper use of public funding, the ethicality of specific practices, or the severity of punishment. These types of issues are often the topic of political debate and are most likely to create strong and differing opinions

making the effects of normative influence decidedly applicable to Civil Good. Since normative influence is defined as a pressure to fit in with the majority [183], it is safe to assume that this pressure would be reduced, if not removed, in a one-on-one debate where there is no immediate group majority.

The advantages of privacy on the Internet also contribute to a stronger minority voice. According to the Asch-style conformity experiments discussed in Section 6.2.1, people exhibit a tendency to conform when they observe that majority opinion may be different from their own even if that opinion is objectively incorrect [9]. It was observed that when people were required to publicly announce their opinion, they conformed significantly more than when they were able to privately note their opinion. In Civil Good, where privacy is a top priority, users should feel less pressure to conform with the majority opinion.

### **6.2.3 Conclusion**

Civil Good is among the first to provide a voice for the minority opinion to this extent. As explained here and in Section 4.2, discussions in the form of one-on-one debate improves interpersonal attitudes, reduces group pressure to conform, and boosts overall conversation quality. By offering a place where they will not be persecuted for their opinions, the service becomes an attractive alternative to in-person debate or forum-style debate for the politically unpopular. At least within Civil Good, the spiral of silence should be mitigated.

## **6.3 History and Political Climate**

At present, political climate is, on the whole, determined through analysis of public opinion surveys [47]. While surveys have proven to be fairly accurate indicators of political climate over the last several decades, they are not without their limitations.

Civil Good could provide a fresh look at political climate, historical or otherwise, by offering a more personal perspective of political climate tied with the demographic information of its users.

### **6.3.1 Background**

Traditional surveys suffer from some degree of response bias and, in some cases, response falsification [122]. Response bias is broadly defined as the conditions during the responding of a survey that negatively affect the accuracy of the responses [226]. Types of response bias range between (1) the motivation to provide “correct” survey responses to (2) response falsification for personal benefit [122]. While this bias is not always significant enough to invalidate a survey’s results, the effect is always present and results must compensate for response bias.

Surveys are also only effective if they are conducted using a sampling that is representative of the entire population. Telephone surveys, currently the most popular type of public opinion survey, are often used due to the capability of random dialing [47]. They take advantage of random sampling which reduces the possibility of having a respondent sample that comes from a narrow demographic. This is effective because neither the respondent nor the survey issuer has any control over who gets chosen. Unfortunately, telephone surveys cannot completely represent the entire population as not all members of the general public own a telephone. In addition, most participants who do own a telephone will be unavailable during their work hours.

Cell phone surveys may be able to address the issue of unavailability especially since landline telephones have become less common overtime [240]. However, they do not completely address the unavailability issue and also come with several other issues. In a paper comparing cell phone surveys with landline telephone

surveys, Michael Link and associates discuss several disadvantages with attempting to reach cell phone-only populations using the traditional style of surveying. Cell phone surveys have a lower rate of participation as recipients of the survey may be busy and unable to participate [240]. Be it while driving, shopping, or working, participants may either fail to respond or provide an inaccurate response due to their multitasking.

Automated telephone surveys (known as “robo-polling”) act as an alternative to traditional live-operator telephone surveys. While cheaper than both landline and cell phone surveys, robo-polling introduces a whole new set of concerns. According to Van Lohuizen and Samohyl’s article comparing robo-polling to traditional surveys, robo-polling suffers from a lower participation rate that they attribute to non-response bias [415]. Specifically, they found a 9% response rate when using robo-polling compared to a 23% response rate when using traditional polling. One of their explanations for this is that each collection method ultimately reaches a different audience (e.g. people who choose to stay on the phone for a live operator versus an automated operator.) Data collection method has a major influence on results and Van Lohuizen and Samohyl stress that results coming from different collection methods should not be directly compared [415].

More relevant to Civil Good is Kevin Wright’s publication on looking at the advantages and disadvantages of online surveys [443]. In comparison to traditional surveys, online surveys have the advantage of being automated, being easily able to reach a wide audience, and being time and cost effective. Online surveys have the disadvantage of lacking participant demographic information, lacking the ability to prevent multiple responses from the same person, and having greater self-selection bias in comparison to live-operator telephone polls. Other disadvantages pertain to possible coverage bias or selection error with online surveys; they can only reach

people who have Internet access and not all who are invited respond [106].

Straying away from traditional methods, the measurement of political climate through the use of Internet resources was examined over a two year period between 2008 and 2009. Brendan O'Connor et al. conducted a study on the reliability of Twitter messages as an indicator of public opinion [295]. They analyzed approximately 1 billion Twitter messages and categorized them based on keywords pertaining to specific topics and the positivity/negativity of the message. Using that information, they were able to piece together public opinion information on topics of presidential approval and consumer confidence over the two year period [295]. Analysis on the validity of the method was encouraging when compared to results gathered from other reliable public opinion polls.

### **6.3.2 Discussion**

Civil Good will be able to provide, at minimum, the non-private conversation logs of its users. On another level, Civil Good could provide information on political climate in a similar fashion to O'Connor's work discussed in Section 6.3.1. Given that Civil Good becomes a service that flourishes over a number of years, it can act as a historical log of political climate over specific time periods or political events. This information is especially useful if Civil Good is able to combine it with demographic information without violating individual privacy.

Using Civil Good as a source of longitudinal political data also reduces the response bias discussed in Section 6.3.1 as the service is geared toward giving users an environment where they are not hindered in their ability to speak their true opinions. In this environment, there is no audience to please and little reason to lie as nothing in Civil Good can be connected with the outside user. However, there is a different type of bias in this scenario: Civil Good does not choose its community;

its users choose the Civil Good community.

Civil Good also offers ways to overcome the disadvantages of the surveying methods discussed in Section 6.3.1. Civil Good addresses two of the disadvantages of regular online surveys: lack of demographic information and multiple responses from the same person. Particularly, Civil Good has more than just survey information that can be used as data (e.g. conversation logs) so the issue of having multiple responses from the same person does not apply.

Additionally, Civil Good has an advantage over surveys in general in the same way that Twitter does in O'Connor's novel method of data collection discussed in Section 6.3.1. Members of both services provide data at their own leisure in the form of text conversation. They are not affected by other activities that interfere with the survey results from landline and cell phone surveys since they provide the data on their own time.

The capability to provide information on historical political climate makes the assumption that Civil Good will become a tool that is widely used among all members of the general population. The wider the audience of Civil Good is, the greater the accuracy and effectiveness of this information.

One limitation of O'Connor's political climate measurement that should be considered is that his results do not cover the gray areas of political opinion. In the case of presidential approval, they were only able to determine whether an individual message approved or criticized the president's actions. Civil Good provides, by a large margin, more information than that of Twitter's messaging system. This information can be used to determine whether or not specific groups hold opinions in between the two extremes.

### **6.3.3 Conclusion**

As a historical log of political climate, Civil Good has the potential to excel. With low cost and possibly publicly available information, the service could become an alternative to the more expensive traditional methods of political climate measurement. Civil Good’s unique functionality helps boost this feature as a valuable tool for political scholars and the like, and such a feature should be considered for Civil Good’s final product.

## **6.4 TED Talks Model**

TED Talks were initially an attempt at sharing the events of the TED conferences with the general public through an audio and video podcast series [397]. The series itself rapidly gained popularity worldwide which was instrumental to the coverage of the original TED conference [397]. Today, both the conference and TED Talks website remain hugely successful. The possibility of utilizing a similar model for Civil Good is discussed in this section.

### **6.4.1 Background and Discussion**

TED’s website is currently ranked 736th on Alexa.com [16] and its online talks have garnered over a billion views [399]. One possible function of Civil Good might be to follow a similar “live event model” [204]. Ideally, if the Civil Good community’s passion for debate is akin to the TED conference’s passion for provoking thought and change, then it may be practical to organize live debates in front of an audience, or perhaps debate conferences in a manner similar to what TED has done.

It may be even more practical for Civil Good to start as a conference or live event. TED Talks began with a live conference and not with a website [397].



Early on in TED's creation, demand for its content was high due to its quality and exclusivity; only a few years after starting, the conference became profitable [397]. If Civil Good could emulate this quality in a conference, it would bring excellent publicity that would help form the right community of potential users before the website is even up. In an interview discussing Civil Good's potential usability, Professor Soussan Djamasbi of WPI stated that if there are too few users on a site like Civil Good, then potential users may feel less compelled to stay [102]. It would be advantageous if Civil Good were able to draw an audience before its actual launch as a website in order to reach a critical mass of users immediately and be off to a stronger start. It may even be possible to branch off of TED by starting Civil Good as a customized TEDx conference.

It is worth mentioning that online content and live conferences seem to augment and serve as advertisements for each other. For instance, the year TED began putting its content online for free, the conference was able to increase its price by 50% and still sell out of tickets in 12 days making them even more profitable [398]. Providing free TED Talk content online actually increased demand for the live TED conference.

#### **6.4.2 Conclusion**

If Civil Good can operate in a similar paradigm to TED Talks, Civil Good conferences could become a tool for the reduction of political divide and the dissemination of debate skills. Civil Good could also use this model as a source of income to finance the operational costs of the website. This model could serve as a way of not only spreading ideas, but also as a way of spreading awareness of Civil Good itself.

## 7 Infrastructure

We researched the different types of hosting available for websites in general, and examined which type of hosting would work best for a website like Civil Good. We recommended the database type that Civil Good should use after researching the different types and their associated pros and cons. We also researched programming languages and open source technologies that are particularly suited to the nature of Civil Good, and thus could be easily leveraged.

### 7.1 Web Hosting Requirements

Civil Good is a Web-based service and therefore needs to be hosted on a Web server. There are several different options for Web hosting, each with its own advantages and disadvantages. In this section, each of the advantages and disadvantages are analyzed to determine which form of Web hosting is optimal for Civil Good.

#### 7.1.1 Background

There are several common factors to be considered in choosing a method of Web hosting. These include uptime, bandwidth, disk space, support, and cost [159, 199, 319]. Uptime is the percentage of time the server is running and able to provide access to the website. Bandwidth is the amount of allowable data transferred through the server by users of the website. Disk space is the amount of space needed to store information from the site. Support is how much assistance is offered in maintaining the server. Lastly, cost is the price to keep the server running.

In choosing a method for Web hosting, one major decision to make is whether the server for Web hosting is built and managed completely by the owner of the server (in-house hosting) or whether the server is outsourced to a Web hosting

company [288, 145, 115, 189]. The advantage to in-house hosting is full control over the server which allows for a high degree of configurability [288]. However, there are many disadvantages to in-house hosting, such disadvantage is the around-the-clock maintenance needed to maintain the uptime of the server. This either requires the full attention of the server owner or a dedicated IT staff [288, 145, 115, 189]. Another disadvantage is the technical expertise required to keep the server fully running and secure. This requires the owner to be an expert or requires the expertise of a hired IT staff [288, 145, 115, 189]. One final disadvantage is the overall cost. A dedicated high speed Internet connection is needed as well as potentially an IT staff, which is significantly more expensive than outsourcing to a company [288, 145, 115, 189].

Outsourcing the server to a dedicated hosting company can help to mitigate these disadvantages. Hosting companies generally operate on a large scale and so can provide Web hosting to others safely and more economically [145, 115, 189]. Moreover, there are several different types of outsourced Web hosting available depending on the needs of the server owner. These include colocation, dedicated hosting, shared hosting, Virtual Private Servers (VPS), and cloud hosting [137, 114, 257, 202].

The first of these options is colocation. In colocation, a company is hired to manage the housing of the server and the server Internet connection, while the owner is still in charge of all other management of the server [137, 148]. The advantage to this is that a website owner no longer has to worry about managing the space and the connection to the Internet. This helps to alleviate the costs created from in-house hosting and reduces the amount of time and labor required in maintaining the uptime of the server [137, 148]. However, the owner is still responsible for maintenance of the server, which requires time and technological expertise [124, 148].

The second outsourcing option is dedicated hosting. In this case, an entire server from a Web hosting company is leased to a website owner [137, 114, 257, 202].

The hosting company owns the server while the website owner can access the server to install programs for the website [137, 114, 257, 202]. Additionally, the Web hosting company can offer managed and unmanaged maintenance on the server. If the managed maintenance option is chosen, much less time and technological expertise is required from the website owner [137, 114, 257, 202]. Lastly, the disadvantage of dedicated hosting is that it is the most expensive of the outsourced Web hosting options because the website owner must lease an entire server [137, 114, 257, 202].

Another option similar to dedicated hosting is VPS. In this case, a server is provided by a Web hosting company on which several websites are independently hosted [137, 114, 351, 257, 202]. Each hosted website is allocated a piece of the server that acts like its own dedicated server [137, 114, 351, 257, 202]. This generally provides similar advantages to managed dedicated hosting [137, 114, 351, 257, 202]. In addition, since the server is divided it is cheaper in cost [137, 114, 351, 257, 202]. However, this division also causes a division in resources, which means that each site can support less traffic [137, 114, 351, 257].

Similar to VPS in the concept of division of resources, a fourth option is shared hosting. In this case, many websites share the same server provided by a Web hosting company, but instead of the server being divided into parts, like in VPS, websites share the same resources on the server [137, 114, 257, 202]. Due to this, shared hosting is one of the most economical of the Web hosting options [137, 114, 257, 202]. However, this leads to a few disadvantages. One is that website owner has very little control over the software on the server [137, 257]. Another is that one of the other websites can potentially use up more bandwidth and disk space to starve other websites on the same server [137, 257]. Lastly, there are potential security concerns. Since websites share the resources of the server, if one website is compromised by a hacker, that hacker can potentially use those resources

to compromise other websites on the same server [137, 75].

The last option is cloud hosting. In this option, the website is hosted across several different servers from several different locations by a Web hosting company [137, 114, 351, 257]. The main advantage to this is redundancy, which means that on the small chance that one of the servers stops functioning, there are others to back it up to keep the website running [137, 114, 351, 257]. Additionally, it is easy to scale to larger traffic since more servers can be easily added [137, 114, 351, 257]. On top of this, it is an economical approach as long as the traffic to the website does not spike [137, 114, 351, 257]. However, because the hosting for the website is spread out across many servers, there are more points of vulnerability, which creates potential security concerns [137, 351, 257].

Most shared services, such as VPS and shared hosting, “oversell” their server resources because users do not utilize their maximum allocated resources at all times. The practice of overselling relies upon there being sufficient resources to dole out to users as required at any one time, based upon users’ expected usage patterns [129]. In selecting a Web host, it is important to investigate whether, and to what extent, the host oversells its resources.

### **7.1.2 Discussion**

The first item for Civil Good to consider in Web hosting is whether to do it in-house or outsource it to another company. In-house hosting would add additional costs in terms of staffing that Civil Good may wish to avoid. Thus, it is recommended that Civil Good outsources its Web hosting.

Civil Good then needs to choose a method by which to outsource its Web hosting. In order to keep user data private, a major concern for Civil Good is security. Cloud hosting and shared hosting both introduce security risks. As a

result, they are not recommended methods of Web hosting for Civil Good despite their advantages.

This leaves colocation, dedicated hosting, and VPS as options for outsourced Web hosting. From the information provided in Section 7.1.1, the option of VPS is cheaper than dedicated hosting. Additionally, it does not require as much time, labor, and technological expertise to maintain the server as with colocation. This makes VPS a nice middle ground. Civil Good may also initially start with only a moderate demand as described in Section 11.2. This means there will be fewer users to consume bandwidth; therefore, the disadvantage of VPS having less bandwidth for use is acceptable. Lastly, the bandwidth of VPS do not have the potential of resource starvation as in shared hosting. This implies that conversations can still run smoothly. Considering all these factors, it is recommended that Civil Good use VPS for its method of Web hosting. But, as the number of users increases, it may become necessary to move Civil Good to a new host capable of handling more traffic. In this case we recommend to use dedicated hosting because it still does not require as much time, labor, and technological expertise to maintain the server as with colocation.

### **7.1.3 Conclusion**

In conclusion, the best option for Civil Good's Web hosting is VPS. Compared to some of the other methods, it requires less labor and technological expertise. In addition to that, it is one of the less costly options. Lastly, its disadvantage of having less bandwidth is also acceptable since Civil Good should only have moderate traffic. Civil Good will have to monitor its hosting requirements as it grows, to determine if and when to move to an option supporting more traffic, which we recommend to be a dedicated host server.

## 7.2 Databases

There are five different types of database management systems (DBMS): hierarchical, network, relational, object-oriented, and document-oriented [385, 380]. Each type of DBMS has its own strengths, drawbacks, and purposes at which they excel, so it is important to determine which type of database would be best for Civil Good. In addition, it is useful to roughly estimate the amount of resources a database system would utilize.

### 7.2.1 Background

A hierarchical DBMS is structured like a tree, with a root node (or *record*) on top and its subordinate nodes connected below. Related records are linked to each other by branches called *fields* that connect to the record that they are subordinate to, forming a parent-child relationship between higher-level records and lower-level records [385]. Because of this simple, predefined structure, it is easy to locate or correct data in the database. However, the hierarchical structure for databases is one of the oldest data storage systems, and because of that, the structure is very inflexible. While a parent record can have multiple children, a child record can only have one parent record connected to it. Additionally, child records cannot be connected to each other even if they are closely related to each other. Lastly, attempting to add new records or fields to the existing structure requires that the entire hierarchy to be redefined. So while it is easy to navigate and easy to alter data, it is very difficult to alter the structure of the database itself.

A network DBMS is similar in structure to the hierarchical model. However, network model records scatters records (in this case called *members*) about as opposed to the tree-like hierarchical system's more linear structure. Also, members can have multiple "parent" records, called *owners*, in the network system [385]. Due

to these facts, a network DBMS is much more flexible than a hierarchical DBMS. Network databases are still limited in that all their data must still be defined beforehand like in the hierarchical system, otherwise the entire database must be redefined with the addition of new data or connections. Additionally, while members are able to have more connections in the network model, there is still a limited number of connections permitted between members and owners [385].

The third type of DBMS is relational, wherein data is inserted into linear tables as opposed to a hierarchy. In the table, each row, called a record or *tuple*, represents a collection of values, while the various columns (called *attributes*) are the different fields used to interpret the values [119]. Different tables of user information are able to relate to each other by sharing common elements—for example, a table containing people’s names and email addresses can be related to a separate table that contains names, phone numbers, and addresses because the two tables share a common field. Relational databases are extremely flexible and easy to use, especially because in this system, data entries can be added or altered without having to redefine the entire database [385]. Popular relational databases include Oracle, MySQL, Microsoft SQL Server, PostgreSQL, and SQLite [195].

The fourth database management system is an object-oriented one. This system is optimal for handling images, audio, video, and other non-text-based data formats. It does so by storing various media types in reusable multimedia *objects* along with instructions for how to interpret or manage the data [385]. In object-oriented databases, objects are able to possess multiple atomic types, or even other objects, as data attributes [294]. Also, any objects that are subordinate to a particular object inherit that object’s data attributes [49]. Because of this, the code for objects is easily reused and new objects can be easily created and added to the database [49]. One downside of using an object-oriented approach is that because



objects can easily interact and share data classes, creating or modifying one class in the database will result in having to change any and all other classes that also interact with objects in that particular class [294]. Additionally, while a pair of tables in a relational database can easily be linked to each other, it is not possible to connect two data classes to one another. Thus, the queries that can be performed in an object-oriented database system are highly limited [294]. The most popular object-oriented databases include Caché, Db4o and Versant Object Database; however, none of these rank very high on the overall list of database management systems [194]

The fifth type of database management system is document-oriented. Instead of describing the attributes of and relationships among the data entities, document-oriented databases focus on organizing information into documents [177]. Document databases can be even more flexible when the data schema changes, and the application can modify the structure of the data at any time [177]. However, this can pose problems in terms of data integrity and consistency. For example, if the application attempts to modify the data structure but only does so for part of the documents, errors may occur when accessing the documents that were not altered. The document-oriented model has the major advantage of encouraging scalability since data for most operations is held in a single document, reducing the need for table joins or multi-object transactions [177]. The most popular document-oriented database systems in use as of 2013 are MongoDB and CouchDB [193].

Among the five database system types discussed, only two are really used in modern Web applications: relational and document-oriented [196]. Document-oriented systems, also called document stores, excel in terms of speed and scalability, and they allow the developers to “quickly and concisely model complex domains,” and provides the most flexibility for quickly-changing data schema [177, 246]. How-

ever, many document stores lack in database transactions, a group of database operations that provide atomic, consistent, isolated, and durable units of work that allow correct recovery in the event of failure [246, 274]. Database transactions become important if multiple systems attempt to utilize the database concurrently. A lack of isolated, atomic transactions could result in erroneous behavior if two different systems access and modify the data simultaneously. Head of research and development at Dell Guy Harrison argues that document databases appeal to developers over relational systems because the “entity-relational data model is usually inherently different from the object-oriented model of modern programming languages” [177]. However, relational models are still more effective in batch processing and have a large set of existing tools available for reporting and data exploration [380]. Relational database systems are also much more popular [196].

Estimating the potential size of the database is important both for estimating cost of operational resources and for choosing a specific database management system. One easy way to estimate this metric is to compare the reported sizes of current popular text-based websites. reddit is a site for discussion and dissemination of information. Similarly to Civil Good, it stores mostly text and has a focus on discussion. reddit currently has 85 million unique users, and for that population, reddit stores 2.4 terabytes, which amounts to only about 28 kilobytes per user [344, 347]. However, reddit’s active user base uses data with peak bandwidth at 924.21 megabits per second [347]. Wikipedia is an online encyclopedia assembled by volunteer contributors. It is also largely text-based, plus some pictures and whatever other media might be needed. Though it is not strictly a discussion site, it has a sense of purpose, a community, and a goal, as Civil Good presumably could have. Wikipedia has 20 million registered users, but only really about 80,000 active users [433]. Altogether, the downloadable English version of Wikipedia is just

about 44 gigabytes for all pages, images and metadata, which makes almost 576 kilobytes per active user—perhaps around a megabyte or two if all the other foreign language versions of Wikipedia are considered [431]. By comparison, Facebook had 100 petabytes of data in August of 2012 and roughly a billion users [324, 142, 76]. That would put Facebook at about 100 megabytes per user, which is far larger than the data stored per user on either reddit or Wikipedia. There is quite a range among the database sizes of these websites, but they still provide workable insights. Database size could depend on a large number of factors; for example, productivity of Wikipedia contributors varies widely with country [432].

### **7.2.2 Discussion**

Considering Civil Good is a website that will have a constantly-changing user base, having to redefine the site’s entire database every time a new user signs up would be highly inconvenient. Also, the inability for records to be subordinate to multiple lists at once would prohibit users from being listed under any other listing than the main user base; precluding certain essential features such as the Fave system. Therefore, the hierarchical DBMS system is not recommended. For the purposes of Civil Good, the network system is more flexible than the hierarchical system. Specifically, the ability for members to be connected to multiple owners would allow users to be listed under each other’s Fave lists in addition to the overall user base, among other things, so it would actually allow Civil Good users to connect to one another. However, the issue of the database having to be redefined whenever a new user is added has yet to be addressed in this model, so the network model is likely still not the best model for Civil Good’s purposes. Regarding the relational database system, since a large number of tables can be made and connected to each other in a relational database, it is very easy to link and compare data, which is ideal

for a website like Civil Good. While a very useful system for managing multiple formats of data, the main drawback for an object-oriented database system is that it is very expensive. Considering Civil Good is more of a text-based discussion website than a multimedia-handling one, it most likely would not get much use out of an object-oriented database, so this option is not very well-suited to it. If Civil Good becomes very popular and starts to demand more resources, a document-oriented database system could work better due to its scalability and flexibility of data structures. However, the lack of database transactions is a huge drawback in terms of application stability; a relational database management system would likely be more durable. Overall, since the relational database system is much more flexible than the hierarchical and network models and does not have the predefined data requisite, a relational database seems like the best fit for Civil Good.

It is not necessary for Civil Good to restrict itself to only one type of database system. Relational databases can be used in concert with document stores, taking advantage of each type's strengths. For example, data that is easily modeled in terms of entities and relationships such as users and faves could be stored in a relational database while more immutable, document-oriented information like conversation content and poll results can be saved in a document store. Utilizing these systems together takes advantage of both a relational database's durability and a document store's speed and scalability. Unfortunately, determining which aspects of Civil Good would be best modeled under what type of database system is beyond the scope of this document.

Until Civil Good gathers millions of users, its database size will likely remain relatively small due to its text-based nature. The fact that reddit can keep their database at around 28 kilobytes per person highlights the small data sizes required for sites that are mostly text-based with minimal user data collected. Be-

cause Wikipedia is less-strictly text-based compared to reddit, one might presume that the amount of data stored per user is larger. This appears to be the case, with the difference being slightly over an order of magnitude more data per user. Facebook has far more data stored per user than either reddit or Wikipedia. This is presumably because Facebook is a different type of site in that it collects all the data it can about users and holds not only text but photos and videos that users upload, and stores all of them. Civil Good, if successful, is far more likely to have data requirements per user similar in magnitude to that of Wikipedia or reddit, simply from the text-based nature of the user's contributions, and also because unlike Facebook, Civil Good does not seek to collect and store private user information. Civil Good would likely only require at most a couple hundred kilobytes per user.

### **7.2.3 Conclusion**

In conclusion, a relational database system is probably the best database management system for Civil Good to use due to its high degree of flexibility compared to the hierarchical and network systems, as well as the lack of necessity for Civil Good to go to great expense just to store multiple data types. However, it is also possible to utilize both a relational system and a document store provided that each system is used for its advantages.

## **7.3 Programming Languages and Open Source Technology**

There are several options to consider for the task of designing the Civil Good website. There exist numerous database engines, Web servers, client-side and server-side programming languages, Web frameworks, and other open-source technologies for designing websites. It is important to evaluate the advantages and disadvantages of each of these in order to determine the optimal set of technologies with which to

develop the Civil Good website.

### **7.3.1 Database Engines: MySQL vs. PostgreSQL**

There are two main database engines to choose from: MySQL and PostgreSQL (also known as Postgres). It is generally accepted that MySQL is the faster of the two, being capable of handling hundreds of clients accessing multiple databases simultaneously [18, 275, 151]. However, Postgres has a wider variety of features with superior performance to those of MySQL [18, 275, 121]. MySQL is widespread and popular, so it has a great deal of compatibility and commercial support [18, 275, 151]. MySQL is also well-documented, has superior capabilities for replicating data compared to Postgres, and has several potent storage models [18, 151]. However, Postgres has better security, developer support, and management ease-of-use, and has the same Oracle database compatibility that MySQL has [275, 121]. Postgres also follows ISO (International Organization for Standardization) and ACID (Atomicity, Consistency, Isolation, and Durability) standards, allowing for increased compatibility, while MySQL does not, although MySQL has storage models such as InnoDB that do comply with these standards [18, 275, 151]. Overall, while MySQL and Postgres each have their pros and cons, they are equally viable for the task of developing the Civil Good website. The decision between the two ultimately depends on whether Postgres' wider array of features and superior functionalities or MySQL's greater speed and widespread usage and compatibility are more appealing.

### **7.3.2 Web Servers: Apache vs. Nginx**

Regarding Web servers, the two that will be considered are Apache and Nginx. Apache is the more popular of the two, being touted as stable, secure, and easily configurable [349, 192]. Apache is also better documented and supports a wider

range of operating systems [192]. However, Nginx has comparable functionality, simpler configuration, and superior performance and efficiency [349]. For example, Nginx is able to serve more requests per second. Apache was found to use more memory and CPU than Nginx and still serve fewer requests in the same amount of time [349]. This is because Apache creates separate processes for each request. Nginx uses event-driven, asynchronous, non-blocking architecture so that it can run multiple processes at once using fewer resources and without those processes blocking each other [192]. As a result, Nginx consumes very little memory while processing static pages. Nginx's performance also does not entirely depend on the present hardware, memory, and CPU, as is the case with Apache [349, 192]. Nginx has performance and efficiency on par with Apache, even surpassing it in some areas. Moreover, Nginx's ability to serve multiple simultaneous database requests without consuming large amounts of memory would be highly useful for a site like Civil Good that aims to have a large user base and that will be performing large amounts of database activity.

### **7.3.3 Client-Side Programming Languages**

Almost all websites utilize the same set of client-side languages. Organizations such as the W3C exist to create standards for languages that all Web browsers need to implement [422]. Consequently, there are few options when choosing client-side languages. These languages include:

- HTML—HyperText Markup Language, the fundamental language used to describe the contents of a page on the Web [427, 424].
- CSS—Cascading Style Sheets, a language used to specify how documents are presented to users [428, 429]. CSS allows developers to avoid duplication of

code and store style information common among thousands of pages into a few files [429].

- JavaScript—a scripting language that allows you to dynamically modify the content of the page and asynchronously communicate with the server without needing to reload the page [31].
- ActionScript - the scripting language originally designed to create animations for display on Web pages and is now largely utilized as a competitor to HTML, CSS and JavaScript for building dynamic, interactive sites [278].

HTML and CSS are markup languages that are needed for every website. However, JavaScript, Java and ActionScript are languages that add interactivity and make websites more dynamic. JavaScript is used by 88.9% of all websites, whereas ActionScript (the programming language for Adobe Flash) is only used in 15.9% of websites [426]. ActionScript has the disadvantage of requiring certain users to install a third-party plugin into their browsers, and some platforms, such as the Apple iPhone, officially do not support running Flash on the client side [57]. Development with Flash takes longer, and rendering is often much slower than HTML equivalents [278, 220]. Flash also makes information invisible to search engines, breaks usability standards such as the back button and increasing font sizes, has inconsistent and limited cross-platform support, and breaks search engine site previews [57]. In addition, users often disable Flash in their browsers to avoid seeing annoying advertisements [220]. It is recommended that Civil Good only utilize HTML, CSS and JavaScript for its client-side languages.



### 7.3.4 Server-Side Programming Languages

For server-side programming, PHP, ASP.NET, Java, ColdFusion Markup Language, Perl, Ruby, and Python are among the most commonly used languages [425]. Descriptions of the languages are listed as follows:

- PHP—a general-purpose scripting language that is especially suited for Web development. It is relatively easy to learn and has a large community for support [322, 388]. However, the fact that it is an interpreted and dynamically typed language that is put together from parts of other languages leads to potential security and performance concerns [388, 283]. If there is an error, rather than halt the process, PHP will automatically convert the error-inducing section of code into something that will run, without alerting the user [283]. PHP will prioritize performing an unintended process over aborting due to an error. One security issue with PHP is when it moves data from a location written in one language to a location written in another, resulting in the functionality of the data changing dramatically [283]. Also, version 5.4 of PHP is highly vulnerable to DoS attacks, due to the fact that it will attempt to allocate as much memory as is specified by the “Content-Length” header; a header that anybody can set to any value they want [283].
- ASP.NET—ASP is a tool from Microsoft included in the .NET framework that provides a high degree of functionality for programming large Web applications. It can be written in several languages but is commonly done in C# [322, 388]. The disadvantage to using this tool is that it is complex and thus has a high learning curve [322].
- Java—a more powerful, object-oriented language generally used for higher traffic websites. Server-side code is generally written in JSP. JSP is meant to

simplify programming by reducing the amount of lines of code that need to be written [322, 388].

- ColdFusion Markup Language—a high level tool from Adobe Systems Incorporated. It is designed to be easy to use and to enable rapid Web development [322, 388].
- Perl—an open-source programming language that is good for Web development due to its text manipulation abilities and fast development cycle [388].
- Ruby—an open-source Web programming language. It is made to be easy to learn and to enable rapid Web development [322, 388].
- Python—an open-source, high-level, object-oriented language that can be used for Web programming [322, 388] It is designed for readability and improves productivity [388].

### 7.3.5 Web Frameworks

Node.js is a server-side JavaScript framework. It uses a single-threaded, asynchronous model to support the performance of long-running server processes while consuming very little memory [405].

There are two Ruby-based frameworks to consider: Ruby on Rails and Sinatra. Ruby on Rails uses a model-view-controller (MVC) architecture, which separates code based on its purpose [29]. It is capable of establishing several framework elements automatically without the need for configuration or repetition, and only requires developers to specify exceptions rather than conventions [29]. Sinatra is a framework that deals mostly with creating Web applications and managing HTTP requests [207]. Of the two, Sinatra is better for smaller websites and Web applications, while Ruby on Rails is the superior framework for larger sites and

projects [207]. For situations where either framework is applicable, some developers prefer Sinatra's freedom of design compared to Ruby on Rails' automation of several design steps [207]. Ruby on Rails is the only one of the two that uses a MVC architecture, so as a result Sinatra would not be able to support Civil Good's more complicated processes as well as Ruby on Rails would. A site like Civil Good that has large aspirations would benefit more from Ruby on Rails than it would from Sinatra.

Web.py is a Python framework described as easy to learn but highly "minimalist" [169]. Web.py has few tools to help developers, instead requiring them to build a website from scratch [169]. A far more popular Python framework is Django. The Django framework comes with numerous Web design tools already installed, and its default security aspects are capable of preventing most common Web-based attacks [168, 306]. Django essentially comes ready-made for Web development. Other commonly-used Python frameworks include Pyramid, Bottle, and Flask. Pyramid is a useful framework for advanced users. Unlike Django, Pyramid does not come with all of the necessary tools for Web design, so its functionality depends for the most part on what modules the user has decided to install and add onto it [306, 219]. While it lacks immediate tools and options compared to Django, Pyramid has more in the way of design freedom and overall speed [219]. Bottle and Flask are both extremely simple frameworks, with Flask only being a single-page micro framework [306, 219]. Of the two, Flask runs much faster and is worthwhile as a learning tool for beginners [306, 219]. Among Python frameworks, Django would be the best choice for a site like Civil Good due to having a fine balance between ease of use and a wide feature-set.

CakePHP and CodeIgniter are also two PHP frameworks worth considering. CakePHP is designed to be simple, to create strong Web applications, and to be

easy on developers due to a lack of need for configuration [317, 339]. CodeIgniter is designed similarly to CakePHP, but it addresses several of the complaints that users have about CakePHP. CodeIgniter is well-documented, while CakePHP is highly unorganized. CakePHP automates certain steps in ways that beginning users may not understand. In contrast, CodeIgniter's logic is simple and straightforward, and it only automates the steps that the user wants it to [378, 339]. CakePHP's heavy automation harms its speed, which is why CodeIgniter is by far the faster of the two [339].

Of the frameworks examined, Node.js consumes the least amount of memory to perform its functions, followed by Bottle, Web.py, Django, Flask, Sinatra, Pyramid, and lastly Ruby on Rails with the most amount of memory consumption [377].

When comparing Ruby on Rails, Django, and CakePHP, Ruby on Rails has less freedom of design and is the least popular of the three frameworks, but has better data management and by far the best testing framework and tools [317]. Django has the most freedom of design, the simplest template syntax, and the best content management [317]. CakePHP has higher ease of use and is the most popular of the three frameworks, but it has weak and unsafe data management, less-advanced object relational mapping, and little to no means of performing tests [317]. Of these three, either Ruby on Rails or Django would be suitable for the purpose of designing Civil Good. CakePHP is outclassed by these two frameworks in nearly every way due to its poor organization, questionable logic, lacking object-relational mapping, and heavy automation. Conversely, Django and Ruby on Rails are both well-designed frameworks whose functionalities would prove beneficial to the development of Civil Good.

### 7.3.6 Miscellaneous Open-Source Technologies

There are other open-source technologies out there that are useful for Web design. Aptana Studio is a code editing software tool that is compatible with many programming languages, including HTML, CSS, JavaScript, PHP, and Ruby/Ruby on Rails [54]. Aptana Studio is capable of correcting code in order to fit a certain format, and is capable of determining various browser compatibilities with the code a user is trying to write [164, 54]. Twitter Bootstrap is an extremely popular Web design toolkit. Bootstrap contains a large number of predefined website templates with strong standards to help users immediately begin designing their websites [74, 236]. Bootstrap also features cross-compatibility with a number of languages and platforms [236]. Bootstrap sacrifices flexibility and freedom of design to allow the creation of functional websites with minimal editing [74]. Overall, it would be worth looking into these two open-source technologies in order to aid in Civil Good's development.

## 8 Incremental Refinement and Data-Driven Development

Regardless of the quantity of research that goes into a design concept for a site, once it is running there will still be unforeseen developments, problems that need fixing, and improvements to implement in a timely fashion in order to stay up to date. These changes must be made on a budget while keeping the system familiar to its users. Iterative website design, the process of making small, recurring changes to a website, offers a solution to these issues. However, it is best if fewer problems occur and fewer developments are unforeseen. Data collection offers a way to understand user behavior, anticipate problems, and check whether changes made to the site function as intended. Data allows development to be guided by relevant statistics.

### 8.1 Background

The Internet is a highly dynamic and constantly changing system, so it is relatively safe to assume that websites themselves are also dynamic entities. For example, reddit is constantly under development; it has a development blog where developers post updates about new changes to the site [346]. “Archive.org” creates and archives snapshots of popular websites and allows its users to view changes over time in those websites [24]. On “Archive.org,” one can see how Wikipedia, YouTube, and Google have changed over time as [22, 23, 25]. Facebook also occasionally undergoes changes, but in a less advisable way; Facebook tends to make drastic changes which appear to anger its users [21, 430, 203]. Lampe et al. found that users generally have a positive view of Facebook, but found that that view depends on the functionality of the site [222]. When Digg introduced drastic changes in Digg V4, Digg users reacted negatively and expressed their dissatisfaction by moving to competitor sites

like reddit [144, 260]. After that, Digg burned through investor capital so fast that the company had to lay off 37% of its staff [26]. Websites change over time, and the majority of them are able to balance renovation with user perception.

“Incremental refinement” refers to iterative, frequent, and small improvements to a website over time. Jesse Mutzebaugh mentioned that “incremental redesign,” as he called it, is easier to implement and debug from a programming standpoint; it is easier on a budget, and the changes are less confusing for users [284]. It should be noted that incremental refinement is not occasional, abrupt redesign, which is a similar but problematic alternative. Professor Soussan Djamasbi, user experience expert at the Worcester Polytechnic Institute (WPI) School of Business, mentioned that for sites with a relatively complicated interface, it is best not to change the interface frequently so that users are given time to familiarize themselves through their investment of time in the site [102]. This could explain why when Facebook abruptly added the Timeline and the News Feed, it outright angered some users [222, 21, 430, 203]. Facebook’s dilemma also seems to agree with Zoltán Góczy’s statements that changes incur a switching cost on users, and that abrupt change is disliked even if the changes are good in the long run [152]. It is also argued that large changes are more costly and risky, and that they leave a site perpetually falling behind the performance of the rest of the Web [152, 354, 160]. Incremental refinement is also not simply adding more features to a site over time, but rather improving or replacing features of a website over time. As Antoine De Saint-Exupery put it, “You know you have achieved perfection in design, not when you have nothing more to add, but when you have nothing more to take away” [125]. Research has demonstrated that simplicity and ease of use are key to Web design; the mere addition of functionality goes against this by gradually complicating the site [37, 56]. Professor Djamasbi also specifically recommended against clutter [102].

Incremental refinement is a process of making small but frequent changes that for the purposes of Civil Good do not include addition of more functionality. Whether these changes are improvements can be found by applying data-driven development.

Data-driven development refers to website development guided by the collection and analysis of data. Data-driven development can be thought of as having three key parts: a method of data collection, a framework or context for analysis and decision-making, and a plan of action.

When it comes to choosing a method of data collection, there are plenty of options for Civil Good. One of these options is surveying—asking users or target demographics about what they want. In-person surveys with non-monetary incentives have the advantage of improving response rate and response quality without having a significant effect on measurement error due to the incentive [436]. Email surveys and mail surveys are another set of alternatives. Some research has been done with incentivized mail surveys, but though incentives improve response rate, they do not significantly affect response quality and may end up creating unintended biases. These biases drive researchers to look for alternative methods of improving response rates [376, 27].

Mechanical Turk offers yet another alternative surveying method. Mechanical Turk is a marketplace where quick but not-easily-automated tasks are performed by humans, called “Turk Workers,” for a small fee. Mechanical Turk has a diverse demographic that is roughly generalizable to the population of the United States and is becoming more international [146, 11]. On Mechanical Turk, though survey participation is hindered by low payment for tasks, response quality is not damaged and the data obtained is at least as reliable as data obtained via other methods [271]. Maintenance of a positive reputation with the Turk Workers is recommended to ensure a good response rate and response quality from the site. It should be kept in



mind that there is some system gaming, though not enough to make data any more unreliable than data obtained via other methods, and combatting system gaming seems to cause some selection bias [146, 373].

Free online surveying services with additional services for a fee are available at “surveymonkey.com” and “limeservice.com,” as recommended by Karine Joly [206]. These services do their very best to keep user data safe as reflected in their policies [239, 392, 393]. These services are widely used in online surveying, but if one would rather take their user’s data safety into their own hands, it is also possible to set up surveying methods on one’s own site, as put into practice in a case study by George Carole who set up a survey for feedback from library users on that library’s website [150].

Surveying is a significant aspect of data collection, but self-reported behaviors and preferences only go so far. Usability studies carried out by Tolliver et al. show that actual user behavior can be counter-intuitive, and recommend a continual process of collecting and analyzing usability data to account for changes in user behavior over time [407]. Professor Soussan Djamassbi also mentioned that user behavior can be completely different from reported behavior in surveys [102]. There are other methods of data collection that should be used along with surveying to get a better sense of which design decisions actually work. Besides SurveyMonkey and Lime-Service, Karine Joly also recommended Google Analytics and crazyegg.com [206]. Google Analytics is a suite of tools for collecting all manner of data about user behaviors and performance on a site, and analyzing that data [153]. In a case study, Google Analytics was used by Wei Fang to improve a library website [128]. “Crazyegg.com” offers a service that shows where and how frequently users click on a Web page, how far they scroll down a page, and how behavior differs between visitors referred to the page from other pages [80]. Karine Joly further recommends

observing what works and what does not for one's competitors [206].

There are other methods of data collection besides surveys to be considered for use. George Carole discusses a method he used instead of surveying: with consent, tape record users as they spoke their thoughts aloud and used a website [150]. Kohavi et al. detail how to implement controlled experiments on the Web in their article "Controlled Experiments on the Web: Survey and Practical Guide," [217]. Controlled experiments randomly assign different treatments to sample populations and attempt to measure the differences those treatments cause in those populations. The simplest controlled experiment on the Web is the "A/B test," which randomly assigns users of a website slightly different versions of that website to use, and collects behavioral data, site performance data, and anything else deemed relevant, to check if the difference in website version causes a difference in the collected data. There also exist methods implementing "Multi Variable Tests" which implement more than two different possible treatments at a time, or attempt to vary more than one website feature in a given treatment [217]. Kohavi et al. explain that controlled Web experiments can be used to find causal relationships between site features and user behavior. Kohavi et al. and Crook et al. [217, 81], caution against small sample size, confounding variables like variation in user behavior over time, and statistical limitations. Tyagi et al. demonstrate the functionality of "WebLog Expert," which allows one to pull up the server logs of their site and do simple performance analyses based on the frequency of error messages for various pages [414]. Moreover, Poblete et al. describe a method for mining the text on pages of a website to evaluate how similar they are relative to how much traffic they get [318].

Once data is available for analysis, the right methods must be used to ensure that the data is properly interpreted and used to motivate the proper course of action. Some general suggestions from John Elder about what not to do with data

mining can be summed up as “make sure to have the right data, know what to ask and have some background knowledge, and do not trust data just because it makes sense while distrusting data that does not” [116]. One frequently emphasized method is learning about the users and their behaviors, and prioritizing the improvement of the features, content, and pages that users actually use and search for [354, 128, 65, 94]. The general idea is that on the Web, just a fraction of content gets the majority of traffic, and the vast remainder gets a small amount. With that in mind, if the pages, features, and content that people actually use are improved, then for the least expense, difficulty, and deviation from the original site structure of the website, the most amount of user satisfaction can be generated.

The technique described by Poblete et al. both mines data and provides a basis for its interpretation. It analyzes the text and traffic of various pages on a website and suggests which pages should be linked together because they are similar, and which pages are more important to link properly with other pages because they get more traffic, and will therefore be more visible. The reasoning behind this technique is that usability and navigability of the site can be most effectively improved by making sure all the most popular content and pages are logically interconnected [318]. A similar approach by Kitajima et al. used Markov chains to evaluate usability quantitatively by emulating how a person would browse the site to find information [214]. Markov chains involve matrices which represent the probabilities that given one state, something will switch to another state. In a Markov chain model of Web browsing, given that the user is on one webpage of a website, there is some probability that they will move to another given page of that site website. Estimating these probabilities and then “walking through” the browsing process can help evaluate whether a site has problems with navigability.

Controlled experiments allow for statistical analysis which produces excellent

resources for decision making. However, much caution must be taken to ensure that ability. Crook et al. make seven recommendations against accidental misuse of statistical methods, and for the control of extraneous variables, like filtering out Web crawlers [81]. Their recommendations attempt to address common problems facing controlled experiments. Kohavi et al. cover many of the same issues in their guide to controlled experiments on the Web [217]. They mention that quantitative metrics are not explanations, and the newness of a tested feature creates bias. They also conclude that short-term effects during the study may be measured, but long-term effects may go unnoticed. Users also tend to become biased when they notice that certain pages they access appear differently when accessed by their peers. Kovahi et al. also warn that studies done in parallel may pollute each other's data.

Another, less structured tactic for interpreting data is heuristics: essentially just human decision-making. Tolliver et al., George Carole, and Wei Fang demonstrated the use of heuristics when interpreting the data to decide on changes to their respective sites [407, 150, 128]. However, caution must be taken with heuristic methods: people are not perfect, they make mistakes. Tolliver et al. dealt with this by working with a consultant experienced with usability studies, George Carole used an iterative technique to check whether the assumptions made in redesign were valid, with a prototype before the final design, and Wei Fang used multiple revisions of a prototype passed back and forth amongst peers [407, 150, 128]. Some of these techniques for mitigating the error-proneness of heuristic methods are also recommended elsewhere: Feelders et al. recommend expertise in data collection and data analysis, as well as experience with the broader context [132], and Elaine Chou recommends steps of prototyping and development when redesigning complex websites [65]. There are three broader contexts within which to consider data and make decisions: user-centered design, activity-centered design, and goal-directed design,

as discussed by Ashley Williams [435]. User-centered design attempts to design a website with the user experience as the priority, for example making a website easier and more intuitive to navigate. Activity-centered design takes into consideration how people interact on a website, and tries to foster an environment of more effective communication or collaboration, like a website that allows teams to store and access files and work on them at the same time. Goal-directed design focuses on the goal of the user. If the user wants to learn, an appropriately goal-directed website for that user might have instructional videos and reference guides [435]. These ways of looking at the problem serve to guide heuristic decision-making.

## **8.2 Incremental Refinement**

It is recommended that Civil Good start small and inexpensive, and develop through fast cycles of testing, implementing just small changes each time. This approach has the advantages, mentioned by Jesse Mutzebaugh, of being easier to implement and debug from a Web design perspective, being easy on a budget, and not being as disruptive to users as the alternatives [284]. Iterative development is reminiscent of good programming practice: small changes are made at a time so that if there is a bug, it is easier to correct. Even from a non-programming perspective, changes may negatively affect usability of the site in unexpected ways, and these mistakes can be more easily tracked down and fixed if only a couple lines of code change with each revision. Iterative development also allows mistakes to be fixed sooner and reduces monetary risk since each iteration of development is relatively inexpensive to implement or backtrack on. The alternative of occasional redesign simply does not perform as well, and has caused problems for Digg and Facebook [144, 260, 26, 21, 430, 203]. It is possible the only reason Facebook can make drastic changes without suffering is because it has tremendous social “momentum” and can afford to

anger some small fraction of its billion or so users [324, 142], who will be reluctant to leave and sever their online connections with friends. Each redesign is more expensive, time-consuming, complicated, and difficult for both for Web designers and users, and infrequent redesigns tend to bring a website up to date with the rest of the Web, only to have it fall behind between redesigns [160, 152]. Mistakes also have more time to cause damage before the problem is corrected, if the time between redesigns is large. The alternative of having no redesigns at all is not advisable. As discussed in 8.1, the Web is fundamentally dynamic: at the very least websites need to update to stay secure and compatible. Furthermore, to expect perfect execution of the website on its first design is a nearly unachievable ideal: our understanding of human behavior on the Web is insufficient to make perfect predictions about behavior, and writing flawless code without any bugs ever is difficult. Incremental refinement would allow Civil Good to adapt in the dynamic environment of the Internet for less expense while keeping the system familiar to its users.

### **8.3 Data-Driven Development**

Civil Good would benefit from collecting and analyzing its own data to inform development decisions. The data sets that Civil Good collects, unlike papers found through scholarly research, can be targeted to answer the questions Civil Good poses. Moreover, good data makes it unnecessary to rely on intuition, which as Professor Soussan Djamasbi said, does not always mirror reality [102]. Data Driven Development allows the relevant facts, free from interpretation, to determine the design decisions of Civil Good.

### 8.3.1 Data Collection

Surveys are one way Civil Good can collect the data it needs, but not all survey methods are equal. In-person surveys can get a high percentage of passersby to respond with high response quality, but they simply cannot reach as many people as online surveys and which may only be generalizable to the local community where the survey takes place. For example, as Professor Soussan Djamasbi noted, a survey of members of the student population at WPI is not generalizable to the population of the United States [102]. Assuming that Civil Good will be based in the United States, it would be best to collect samples for surveys representative of the population of the United States. In-person surveys might be capable of this, but only if data is sampled from across the country, which could be time-consuming and expensive. It might be easier to collect data generalizable to the United States with mail surveys or email surveys, but those methods run into issues with non-responsiveness and poor data quality [376, 27]. A good alternative would be surveying on Mechanical Turk, since that has manageable system gaming, the promise of large and diverse samples generalizable to the United States, and a relatively cheap cost to the surveyor for reasonable sample sizes [271, 146, 11]. Mechanical Turk is an excellent option for any future surveys to collect data useful for making design decisions about Civil Good. However, other survey methods have their strengths and need not be entirely discounted, especially since Mechanical Turk has a shifting demographic that has recently become more representative of India, suggesting that there is no guarantee Mechanical Turk will always be generalizable to the population of the United States [11]. Free online surveying capabilities like SurveyMonkey and LimeService are also available [206]. As Professor Soussan Djamasbi suggested, surveys have the notable limitation that people will sometimes say they would like certain features, but not actually use those features [102]. Surveys are useful and

are recommended for data collection for Civil Good, but ideally data collected from surveys should be collected along with other metrics in order to obtain a clearer understanding of what design decisions should be made.

Google Analytics, CrazyEgg, and WebLog Expert each provide useful data for data-driven decision making, whether it be what users look to on a page and where they click, which pages get the most traffic, records of the server logs, or information about where visitors are coming from [206, 414]. These methods of data collection bring the privacy policy and legal field into consideration; however, despite any perceived flaws the value of the data obtained from Google Analytics, CrazyEgg, WebLog Expert, and other programs for data-driven site development is considerable. Another method of data collection more in line with surveys is collecting recordings to better understand the thoughts of users when using the site [150]. This is not quantitative data, but it can still be useful. Controlled experiments like A/B testing on the Web provide, when done properly, numerical data of quality that is high enough so the data can be used to find causal relationships [217]. Controlled experiments are excellent for making determinations with a higher degree of certainty, but they are also more complicated and nuanced to implement [81]. Each of these data collection methods has strengths and weaknesses, and there is no best method, but there is room to experiment and find what works best for Civil Good.

### **8.3.2 Data Application**

To most effectively understand and make use of data collected to help inform the development of Civil Good, a logical basis is needed for understanding it. Considering this, it may seem that just using human intuition and heuristics when exploring the data may not be the best course of action. However, in practice, simple manual interpretation of data is easier, less formal, faster to implement, and often does pro-



duce acceptable solutions [407, 150, 128, 65, 132]. Additionally, there are ways to combat the limitations of heuristic methods, including informing oneself well about the topic to be studied, working with an expert, having multiple iterations of development, and checking occasionally whether the assumptions being made are correct with a more thorough statistical analysis of data. As long as they are implemented properly, controlled experiments provide a trustworthy method to check if heuristic methods led to the correct design decisions. This method is recommended for Civil Good, since ease of implementation can be balanced with data quality to get frequent data collection, and good data.

There are other methods of data analysis to consider as well. One particularly useful basis for improving websites is the idea that most of the site traffic passes through a fraction of the pages on a website, and if it can be determined what features and content users are actually making use of, one can concentrate on improving those specific features [354]. This is also non-invasive: it can be done without actually changing any of the functionality of the site. Another basis for analysis is to look at how similar pages are and how much traffic they get, and determine based on these factors whether linking similar pages is valuable or not [318]. This method helps improve navigability of the site by prioritizing which pages should be linked together. There is also a technique which uses Markov chains to walk through the task of browsing [214]. This technique also helps improve usability and navigability.

### **8.3.3 Miscellaneous Considerations**

WebLog Expert and CrazyEgg have free 30 day trials and then have paid versions [80, 414]. As one-time expenses for the ability to collect data useful for data driven development, these purchases may be worthwhile to Civil Good. Google Analytics is free to use below a certain level of Web traffic [128]. The actual changes

required by incremental refinement of the Civil Good website will likely have some nonzero time commitment associated with them. If one learns how their site works and how to program it, then they could implement iterative changes themselves - especially if any given change is not too complicated as Zoltán Góczy suggested [152]. Additionally, it is possible that Civil Good enthusiasts with aptitude for programming could donate time for incremental refinement. There is a tradeoff to be made between the time and monetary costs of iterative development, and the benefits it offers to Civil Good.

The process of data collection, analysis, and refinement described in this section works best once Civil Good is running. As David Kelley said: “Enlightened trial and error outperforms the planning of flawless execution” [217]. Data collected while the site is running gives live feedback to how the site is doing, and though surveys can be conducted before the site is running, these cannot effectively collect user behavior data and traffic statistics. If the Civil Good concept were refined indefinitely with the intention that the site should be perfect before implementation, then the site would be “vaporware” - a well-thought-out idea perhaps, but one that is never implemented. Incremental refinement and data-driven development offer a methodology for developing a site from some imperfect prototype, to help avoid that pitfall.

Considering Civil Good’s aim of protecting user information, data-driven development on Civil Good should not pose a privacy risk to the users. It helps that it is unnecessary to keep data used for site improvement for much longer than it takes to analyze the data and come to conclusions. This is because site performance and user behavior vary over time and with changes to the site; there is no guarantee that old data will match the behavior of a new and improved site [217, 81]. For these reasons it is recommended to collect data as an ongoing process [407].

## 8.4 Conclusion

Incremental refinement of Civil Good from a prototype website to a streamlined hub of debate is less risky, less expensive, and easier to follow both on the developer side and on the user side, and more natural on the Web. Data-driven development complements incremental refinement by guiding each refinement of the site and allowing for more informed development decisions. To implement data-driven development, a method of gathering data is required; there are a variety to choose from, including: in-person surveys with non-monetary incentives, Mechanical Turk, SurveyMonkey, LimeService, Google Analytics, CrazyEgg, WebLog Expert, “A/B Testing,” and “Multi Variable Testing.” There are also many ways to interpret data including: heuristic methods, statistical analysis, developing within the framework of “goal-centered,” “activity-centered,” and “user-centered” design, focusing on popular pages and features, focusing on navigability between popular pages, and simulating user browsing with a Markov chain model.

With each way there are pros and cons. There are some monetary considerations and privacy concerns involved here, but these are not sufficient to abandon the implementation of Civil Good through iterative, data-driven development.

## 9 Security

For a website like Civil Good, which aims to preserve user anonymity, it is of the utmost importance that user data is protected. We researched the various methods of ensuring the security of the user experience, and made recommendations on how Civil Good could incorporate security into its framework from the ground up.

### 9.1 Database Security

Database security is necessary to protect the information that Civil Good users deem to be private. As computers become more powerful with time, encryption alone may not be sufficient to protect user data. Encryption algorithms are especially necessary in case the database is compromised in some way. This section discusses methods and countermeasures for protecting a database system from being breached.

#### 9.1.1 Background

Relational database management systems (RDBMS) often include some sort of access control mechanism that allows for discretionary policies for data authorization [38]. Milicchio and Gehrke write about how to develop distributed systems that interact with each other in a secure manner. In this book, they explain that PostgreSQL (one type of RDBMS) can use SSL encryption for its client authentication [149].

Applications that interact with database systems can be vulnerable to a common attack known as a Structured Query Language (SQL) injection. This is a common issue with database-driven websites that is easily detected and exploited [287]. SQL injections are possible because SQL does not differentiate between input data and command structures, and vulnerable applications do little to ensure separation

between data and code [287, 38]. These attacks have been used to gain customer order and payment information from e-commerce sites, as well as gain direct access to confidential information from high-profile companies and associations, such as Travelocity, FTD.com, Creditcards.com, Tower Records, and RIAA [43, 173].

Popular countermeasures to SQL injection generally include writing application code that utilizes automated systems to sanitize data input. For example, the National Institute of Standards and Technology (NIST) recommends to use automated structured mechanisms to automatically enforce such separation by utilizing escape sequences to sanitize any user input before executing commands to insert user data [287]. Some RDBMS contain features such as prepared statements and stored procedures that allow the programmer to preprocess database queries that only accept pure data input as parameters, meaning that the database system will not accidentally interpret user input data as commands [287, 43]. Researchers at the Georgia Institute of Technology (Georgia Tech) analyzed and compared a number of different methods of neutralizing SQL injection attacks, including defensive coding patterns, positive validation and dynamic detection [172].

Another technique that is easy to implement when writing database-driven applications is input type validation [172]. Oftentimes, SQL injection vulnerabilities arise simply from insufficient input validation [287, 172]. For example, a user entering a U.S. postal code should only be allowed to send a four- or five-digit, numeric values; any values that are either non-numeric or contain more than five digits should be rejected.

It is also recommended that developers avoid returning standard SQL error messages to the attacker if a query fails [43]. This is because error messages can provide an attacker with information about the structure of the application and could make finding vulnerabilities easier. The Georgia Tech researchers determined

that defensive coding in the client application is not sufficient, and that “their application is problematic in practice” [43].

An alternative to sanitizing database input is positive pattern matching. Developers would specify models for legitimate SQL queries and positively match the ones that are valid [172]. For example, researchers for AMNESIA built an automated system that combines static analysis of code and dynamic runtime monitoring to programmatically build models of possible valid SQL queries and reporting or rejecting any queries that do not follow the static models [173]. The advantage of this system is that it runs independently of the application code and serves as a check for any other clients that directly interface with the database system. In addition, positive validation mitigates the risk whereby developers fail to validate input for every possible attack vector. Other non-development techniques include detection and prevention systems such as intrusion detection systems (IDS) and black box penetration testing tools such as WAVES [172]. Systems such as these deal with the problems that other techniques ignore such as poorly-coded stored procedures [172].

### **9.1.2 Discussion**

Civil Good’s database security can benefit from utilizing all of the techniques discussed in Section 9.1.1. To protect the database itself, the database management system should run in a separate server from the application server, communicating only through connections secured with public key encryption such as SSL. The database server should be run separately from the application server, and communication between the two servers should be securely authenticated with SSL. Database permissions should be restricted to the minimum necessary for application functionality. Civil Good should avoid SQL injection by sanitizing inputs before inserting into the database and utilizing software that automatically escapes and filters data.

In addition to sanitizing inputs, AMNESIA can be used to perform positive pattern validation on queries. Civil Good can also use input validation to verify its users are inputting correct data. Additionally, Civil Good should also avoid printing standard SQL errors. Most importantly, Civil Good’s application code should never dynamically construct and execute query strings using user input [287].

### **9.1.3 Conclusion**

In general, database security relies most heavily on controlling access and preventing attackers from exploiting vulnerabilities in the Civil Good application server. To mitigate this risk, it is recommended that Civil Good run the database management system on a separate server from the website. Civil Good should use public key encryption for communication between the database and website. Furthermore, database permissions should be restricted to the minimum necessary for application functionality. Within the application layer, Civil Good should mitigate the risk of SQL injection by sanitizing inputs before inserting them into the database, using input validation to verify data, and utilizing software such as AMNESIA. Most importantly, to prevent SQL injection, Civil Good should avoid dynamically constructing and executing query strings using user input, and printing standard SQL error messages to the end user.

## **9.2 Common Attacks**

Although not every computer exploit is the same, certain vectors of attack can be categorized and grouped together. Some kinds of attacks such as “Confused Deputy,” Cross-Site Scripting (XSS) and Cross-Site Request Forgery (CSRF) take advantage of software vulnerabilities in Web applications to execute unauthorized code through the service [287]. Other common types of attacks, known as Denial of

Service (DoS) attacks prevent a service (such as a website) from being accessed by its intended users. As with any public-facing Internet service, Civil Good needs to be designed and implemented such that the risks of these attacks is minimized.

### 9.2.1 Background

The U.S. Department of Defense (DOD) released a whitepaper that outlines common vulnerabilities in industrial control systems, many of which may relate to Civil Good. One method listed, known as Cross-Site Scripting (XSS), “allows attackers to inject code into Web pages” generated by vulnerable sites, usually caused by “poorly sanitized data” [287]. This method allows attackers to execute their malicious code on user systems rather than having to access the Web server. This can have enormous repercussions, including the disclosure of end user files, the installation of Trojan horse malware, redirection to phishing sites, and modification of content presentation [287].

XSS attacks are easily mitigated through sanitizing or escaping any user-submitted data on output. For example, if a user were to submit a post on Civil Good that contained a snippet of malicious code, the Civil Good server would make sure that the post is encoded such that the malicious code cannot be executed. Web applications achieve this by ensuring that “[a]ll non-alphanumeric client-supplied data are converted to HTML character entities before being redisplayed to a client” [382].

Websites can perform data sanitization either on input user data before it enters the database or on database output before the final HTML page is sent to the user. Rasmus Lerdorf, one of the original creators of PHP (a popular Web programming language) mentioned briefly in a blog post that he uses a PHP extension that “automagically sanitize[s] all user data for [him]” on input, which he claims prevents



XSS attacks [234]. Luke Plant, one of the core developers for the popular Python Web framework Django, criticizes this practice claiming that sanitizing on input is a “transformation that is totally irrelevant to that data” [316]. Plant explains that because the data could potentially be represented as a variety of different media types (each of which requires a different sanitizing method), it is best practice to store the raw, unmodified data and sanitize the data on output instead. Sanitizing on output also has the advantage of being an upgradeable process. Consider, for example, that a new exploit is found for the output medium that was not taken into consideration when writing the input data sanitization code. The application is now vulnerable, and the only way to patch the vulnerability is to retroactively apply the necessary encoding to the existing stored data. If it were designed instead to sanitize on output, the code could simply be updated to patch the new exploit without needing to change the database content.

In a “Confused Deputy” attack, a program is tricked by another party into misusing its authority [383]. Cross-site request forgery (CSRF), an instance of the “Confused Deputy” attack, can occur if a server does not sufficiently check the authenticity of a user request [209]. CSRF attackers hijack credentials and “trick a client into making an unintentional request to the Web server that will be treated as an authentic request” [287]. The term “cross-site” means that the request is sent from the user’s computer to the targeted site, but the actual request comes from a site controlled by the attacker. Attackers can use this method for arbitrary actions on a vulnerable site while assuming the identity of the victim.

CSRF attacks are usually triggered when the victim visits a page controlled by the attacker. The basic premise is that the victim visits the attacker’s site, and the victim’s browser sends a request to the targeted site using whatever authentication tokens or cookies are already present in the victim’s computer. There are

numerous attack vectors including hidden images whose source attribute is set to a vulnerable site's URL and hidden forms on the attacker's page that get automatically submitted with JavaScript [209]. The attacker could send an asynchronous request or exploit vulnerabilities in the victim's installed browser plugins such as Flash or Java [209]. In each implementation, the victim's credentials are hijacked, and the request is sent without the victim's knowledge. Since a vulnerable server cannot tell the difference between a legitimate user request and a cross-site, forged request, the attacker can do anything the victim could do on the vulnerable site.

Nelson and Chaffin of the U.S. Department of Defense recommend that “vetted libraries and frameworks that provide functions for implementing CSRF mitigations” be used [287]. Konstantin Käfer, Web developer for MapBox [253], highlights some other measures for mitigating CSRF attacks including checking HTTP Referer headers and generating CSRF tokens [209].

The simplest defense is to “validate the HTTP Referer header, preventing CSRF by accepting requests only from trusted sources” [35]. The HTTP Referer header is a part of a request that indicates the origin site of the request. For example, clicking on a link to Wikipedia from Civil Good's website places “https://civilgood.org” in the HTTP Referer header. By checking that the Referer header comes from a trusted domain, the application can in theory block all cross-site requests. However, both attackers and genuine users may have clients that send requests that do not have the Referer header. Blocking requests that lack a header could potentially block real users, but allowing requests with missing Referer headers creates a vulnerability that allows attackers to bypass the HTTP Referer check if the victim's browser omits the Referer header.

The most popular CSRF defense is to generate and include a secret token with each request and validate that the token is correctly associated with the user

session [35]. A valid token would indicate that the user elected to submit a request from the original site because a cross-site forged request could not have access to the token. A long enough token should prevent brute force guessing by a CSRF attacker; the recommended length is 15 characters or more [287]. Like HTTP Referer validation, CSRF tokens do have their pitfalls. Although a correct implementation works well for a logged in user, forged login requests cannot be blocked with a token because guest users have no login session with which to bind the CSRF token.

Another common form of attack called Denial of Service (DoS) involves the attacker denying a normal service (such as a website) to legitimate users; it can be launched through a variety of means [62]. Some approaches, such as buffer overflow attacks, take advantage of poorly-written application code to simply bring down the service [287]. Other types of DoS attack use distributed systems to flood a service, usually by using compromised hosts to overload the victim with useless packets such that the amount of traffic exceeds the amount that the service can support and so that normal users cannot access the service [62]. This kind of attack, known as “distributed DoS” or “DDoS,” has been widely publicized since February of 2000, when even well-equipped networks such as those of Yahoo!, eBay, and Amazon.com fell victim to this kind of attack [225].

Non-distributed DoS attacks that exploit application-level vulnerabilities can be mitigated in a similar manner to XSS attacks, that being through exhaustive tests and thorough review of code. Common vulnerabilities noted by the U.S. DOD include improper input validation, improper authentication and insufficient verification of data authenticity, all of which can be mitigated through careful code review [287]. Production applications should always be “updated with the latest security patches and techniques” [225]. Other security measures include disabling unused services and setting up intrusion detection systems [225]. Because of the

large variety of methods that denial of service attacks can leverage, methods such as these can be used as preemptive measures to secure against them.

The distributed nature of DDoS attacks makes application-level measures, such as updating with security patches, ineffective in dealing with them. Since DDoS attacks work by exhausting resources, other pre-emptive measures are needed. Rocky Chang of the Hong Kong Polytechnic University suggests monitoring network traffic for common attack signatures and using packet filtering to drop malicious packets [62]. Another approach is rate-limiting the speed at which packets are received [62]. DDoS attacks have many forms and no single solution exists to prevent all types of attacks.

DoS attacks are sometimes launched because the attacker has gained unauthorized access to the system and has installed malicious software. Attacks such as these can be mitigated through intrusion detection systems (IDSes) and intrusion prevention systems (IPSeS). IDSes and IPSeS utilize either of two main methods of intrusion detection: misuse detection and anomaly detection [171]. In misuse detection, the system monitors the site for a list of signatures of unauthorized or malicious behavior patterns, and alerts for any situation where a user matches one of these signatures [171]. In contrast, anomaly detection defines a list of expected behavior and alerts for any action that is not included in the expected behavior list [171]. These signature matching techniques have high levels of precision and accuracy and are thus widely deployed [17]. Their main weakness is that they can do nothing to prevent against novel attacks (also known as zero-day attacks) [272].

IPS is most effective in environments that manipulate sensitive or confidential information where immediate processing of the data is not critical [171]. The defining difference between an IDS and an IPS is that an IDS is an alert system, but an IPS automatically takes measures to prevent further abuse. For example, IPS SnortSAM

can interact with several different firewalls to add blocking rules for an attacker’s IP address [272]. SnortSAM is an add-on to popular IDS/IPS Snort [289, 272]. Other commercial IDS offerings include Dragon IDS from Enterasys, ISS RealSecure, and Cisco’s IDS [272].

### 9.2.2 Discussion

To protect itself from these attacks, Civil Good’s application code must be written securely and be thoroughly tested for vulnerabilities. The code should be designed from the ground up with risks in mind, and “[s]ecurity should be explicitly at the requirements level” [264]. As mentioned in Section 9.2.1, code peer review is an important part of the development process. Gary McGraw, writer for security consulting firm Cigital, emphasizes risk analysis, external code review, and the formation of a good security test plan during the development of a secure application [264]. By considering security throughout all aspects of the development and maintenance processes, the risk of these common attacks is much more effectively mitigated.

For XSS attacks specifically, Civil Good should assess possible risk areas. Notably, conversation histories and user-submitted poll responses could potentially contain malicious code. As explained in Section 9.2.1, data from these areas should be properly sanitized before sending the page to the user. XSS risks can also be mitigated through design decisions outside of code. For example, Civil Good’s proposed privacy policy in Appendix E prohibits the use of cookies for storing personally identifying information, therefore inherently reducing the risk of a XSS attack being used to disclose such data since the malicious scripts cannot access personal data by stealing information in user cookies. Section 9.1.2 discussed user input sanitization in regard to databases to prevent SQL injection attacks. Data sanitization for XSS attacks is different in that the sanitization process for output depends on the

output medium as mentioned in Section 9.2.1. For example, data accessed as JSON needs to be sanitized differently than data accessed as HTML or XML. If data were sanitized for HTML before being stored into the database, the sanitization would need to be undone if the data were subsequently accessed via a different medium. Therefore, Civil Good should only sanitize for XSS attacks on output.

A CSRF attack could also severely put users at risk. A properly executed attack on Civil Good would enable the attacker to masquerade as a user, post offensive or hateful comments, send bogus seeks or skew the results of polls. As shown in Section 9.2.1 there are numerous methods to mitigate CSRF attacks, but each individual method has significant problems. Barth et al. conclude based on analysis and experimentation that different CSRF defenses work best for certain use cases [35]. For example, login CSRFs can reliably use strict Referer validation since login forms usually submit over HTTPS and browsers only suppress the Referer header over HTTPS between 0.05% and 0.22% of the time [35]. This also applies to other HTTPS traffic. Since HTTPS greatly improves the efficacy of CSRF defense methods, it is highly recommended that Civil Good be only accessible via HTTPS. If Civil Good were to integrate any third-party content (such as advertisements), the site should use an existing framework that correctly implements secret token validation [35].

Among the types mentioned, DoS attacks would likely be the most difficult to prevent because of the high number of attack vectors (and for DDoS attacks, the distributed nature). As with XSS, security and risk analysis for DoS attacks should become an integral part of each step in the software development cycle. Poorly-written code could result not only in denial of service to users, but a loss of trust from community members and the risk of a breach of user privacy. DoS attacks can also be mitigated by installing and configuring an intrusion protection system such

as Snort.

### **9.2.3 Conclusion**

Civil Good’s success in mitigating these common attacks depends largely on multiple facets. It takes a skilled operations team, talented software engineers, and an informed community of users to keep the site secure. To prevent XSS attacks, Civil Good’s application server should sanitize any user-submitted data on output. Civil Good should not sanitize user input for XSS attacks; it should only sanitize on input for SQL attacks. It is highly recommended that Civil Good use HTTPS for all connections. To mitigate the risk of CSRF attacks with the login form, Civil Good should use HTTP referer-watching. Civil Good should also use strong CSRF tokens to validate requests from logged in users. During development, Civil Good should ensure that every line of code is tested and reviewed. The servers should also be configured to rate limit connections to mitigate the effects of DDoS attacks, and should utilize an intrusion detection/prevention system such as SNORT to reduce the risk of unauthorized access and DoS attacks.

## **9.3 Encryption**

Encryption is a technique that uses mathematics to transform information such that it is unreadable to everyone but those with a special piece of knowledge called a “key” [117]. In accordance with Civil Good’s dedication to protecting user privacy, encryption plays a vital role in ensuring that malicious attackers cannot gain unauthorized access to user information. The following sections discuss options for encryption, why it is necessary for Civil Good, and further details about where it should be implemented between end user computers and the server infrastructure.

### 9.3.1 Background

Two commonly used types of encryption include symmetric and public-key (also known as asymmetric) encryption [92]. Symmetric encryption uses a single secret known as a “key” to both encrypt and decrypt the private information [92, 181]. This can be used not only for the exchange of private messages, but also for securing information for storage. Examples of symmetric encryption include block ciphers such as Advanced Encryption Standard (AES) and Data Encryption Standard (DES) [85]. The main issue with using symmetric encryption for messaging is that “sharing a key between the sender and recipient before communicating, while also keeping it secret from others, can be problematic” [92]. Public-key encryption attempts to solve this problem by “defining an algorithm that uses two keys,” a private key and a public key [41, 67, 181]. Its main purpose is to provide privacy and confidentiality [267]. The main disadvantages of public-key encryption are that it does not guarantee data integrity and does not provide that the origin of the data is valid and authenticated [267]. Compared with symmetric encryption, public-key encryption runs more slowly and is therefore most commonly used to transport keys used for symmetric encryption instead [267]. Popular implementations of public-key encryption include RSA and Rabin [267].

Strong encryption methods are often used as defense against man-in-the-middle (MITM) attacks. MITM attacks occur when an attacker reroutes communication between two victim hosts through the attacker’s computer unbeknown to the two victims. This allows the attacker to read and monitor the traffic [273, 95]. This type of attack is very old and has been used against a wide range of protocols, especially login and entity authentication protocols [95]. MITM attacks can even be used to manipulate information in transit between the two victim hosts [444]. Encryption protocols such as IPsec, SSL, TLS and SSH are often used to mitigate



MITM attacks [191]. On the Web, HTTPS uses SSL and TLS, which can be used to secure connections against most MITM attacks. However, it is possible to attack connections secured via HTTPS by exploiting some properties of common local area networks and typical behaviors of inexperienced users [55]. However, these attacks have a very low probability of success.

For most symmetric algorithms, the strength of the encryption comes from the length of the key. Strong encryption algorithms utilize longer keys since they take significantly more time to attack via brute force. However, the tradeoff is that longer keys require more computational resources to encrypt and decrypt the information. AES supports key lengths of either 128, 192 or 256 bits, and Rijndael, a precursor to AES, supports any variable key length whose bit count is a multiple of 32 and at least 128 bits [85]. However, keys longer than 256 bits for algorithms such as AES and Triple DES (an algorithm that applies DES three times) are deemed unnecessarily long and computationally expensive [357, 85]. There exists a technique called “partial encryption,” which works with data compression formats to encode data such that standards-compliant decoders can decompress and render the data without a key at albeit severely degraded quality [357]. This technique is useful for media such as images and videos, and has the advantage of requiring less computational resources than strong encryption algorithms like AES [357].

Kerckhoff’s principle, a fundamental principle in cryptology, states that an adversary who knows all the details of the cryptosystem, including algorithms and their implementations, would have no advantage since the security of the cryptosystem is based entirely on the secrecy of the keys [92]. Delfs and Knebl write that “it is desirable to design cryptosystems that are provably secure” in that mathematical proofs can be used to show that a system can resist certain attack vectors [92]. The ideal is to achieve perfect secrecy, in which attacks where it is impossible for an

attacker who has only the encrypted ciphertext to recover the original plaintext message [92]. However, in most practical encryption algorithms, perfect secrecy is not unconditionally provable, but it is provable that algorithms can come significantly close to perfection [92].

### **9.3.2 Discussion**

Public-key encryption should be used for communication between user browsers and the server. Examples of such protocols are TLS and SSL. As recommended in Section 9.2, HTTPS should be used for all connections between end users and the Civil Good servers since it utilizes TLS.

Strong symmetric encryption should be used to store information in the database. Recommended algorithms include AES and Triple DES [357]. Longer keys provide better security at the cost of performance [85]. Encryption keys need not be longer than 256 bits. Longer keys would be too computationally expensive. The keys should be stored such that compromising the database or application server does not compromise the key. Since Civil Good's media is mostly text-based, partial encryption is likely not useful enough to utilize.

As mentioned in Section 9.3.1 encryption algorithms cannot be proven to be perfectly secure; however, they can be proven to be reasonably close. Thus, it is recommended that Civil Good should do its best to ensure that its cryptographic systems are provably close to secure perfection.

### **9.3.3 Conclusion**

To mitigate the risk of man-in-the-middle attacks for end users, it is again recommended that Civil Good use HTTPS for all connections. Symmetric key encryption algorithms such as AES are sufficient for encrypting data in the database provided

that they have long keys (256 bits or more). Encryption keys should be stored separately from the database, and asymmetric public key encryption should be used for any end-to-end connections between the client and server. All protected information should be encrypted such that its protection is mathematically provable to be as close to perfect secrecy as possible.

## **9.4 Password Security**

Since Civil Good regards the protection of user privacy as a high priority, having a secure method of protecting passwords is of paramount importance. Regardless of the level of security in the backend database, the weakest points protecting users' data are their account passwords. Password security not only involves safely storing and validating user credentials, but also ensuring that the user chooses secure passwords and is informed enough to protect its secrecy.

### **9.4.1 Background**

Passwords are commonly used as one form of authenticating, or positively verifying the identity of, users [300]. IEEE fellow Lawrence O'Gorman writes that authenticators can be characterized as knowledge-based, object-based, or ID-based, each of which is supported by different principles and have different drawbacks [300]. Passwords, a form of knowledge-based authentication, are characterized by secret or obscure information, whereas object-based authenticators rely on a specific physical item such as a key, and ID-based authenticators rely on something unique to the user that is difficult if not impossible to forge, such as a credit card, a license, or biometric data like fingerprints or retinal patterns [300].

One of the drawbacks of using passwords for authentication is that they require the user to precisely remember a particular piece of information. This is not

something that the human mind regularly excels at, so as a result users will tend to make their passwords easy to remember, and by extension easy to guess [311]. For passwords that are not easily memorizable, users will tend to write them down and keep the information nearby, where that information can be easily discovered by others [311]. Furthermore, users are more likely to write down their passwords if they are required by a website or business to change them on a regular basis [311]. Because precise recall is not a strong point of the human mind, people tend to practice unsafe password protection policies for the sake of ease of memorization.

There are three main ways of discovering a password without obtaining it directly from the user: a dictionary attack, a brute force attack and a hybrid attack [391]. Dictionary attacks employ the strategy of using common words and sequences in order to guess a password. Brute force attacks simply attempt to exhaust every possible combination of letters, numbers, and symbols until it finds the password. Hybrid attacks are brute force attacks that also utilize common words and sequences to attack more common passwords first.

Password attacks can be either online or offline. In offline attacks, password cracking is performed on “a system other than the system on which the passwords were stored.” Offline attacks can be performed in a potentially distributed manner, more quickly and efficiently, and without affecting the original authentication system [358]. Online attacks utilize the authentication system directly, and can be partially mitigated by limiting the rate at which passwords can be entered.

A cryptographically secure system of validating and storing credentials becomes ineffectual if the users choose weak passwords to begin with. Thus, it is important to establish policies that encourage strong passwords. According to recommendations by the National Institute of Standards and Technology (NIST), password strength is determined by the length and complexity (i.e. the unpredictability

of the characters) of the password [358]. Longer passwords with highly varied characters tend to be much stronger.

Websites today employ a multitude of different requirements for password policies. In 2004, Dr. Summers and Dr. Bosworth of Columbus State University published a list of recommendations for allegedly good password policies including requiring that “Alpha, number and special characters must be mixed up” and a “Maximum password age of 45-60 days” among others [391]. These may have been considered important when the article was written, but there is little quantitative research to support the efficacy of such policies. Singer et al. in a 2013 article comment that many standards such as these “are based on inconsistent research at best, and in some cases appear to be pulled out of thin air” [375]. Furthermore, Singer et al. note that requiring certain classes of characters fails to guarantee greater entropy (and by extension increased password strength) since this requirement does not prevent multiple users from sharing similar passwords. Philip Inglesant and M. Angela Sasse at University College London observed that users develop certain strategies for generating passwords [187]. Upper limits on password lengths or character sets break these strategies and cause users to try to cope by making their passwords less secure through such practices as writing them down [187].

Some systems require that passwords expire and be changed regularly in order to defend against brute force attacks. For example, Worcester Polytechnic Institute (WPI) requires that passwords be reset at least every 180 days [441]. Such policies prove problematic because they are ineffective against certain types of common password attacks such as keyloggers and backdoors as cited in a NIST password management guide [358]. Furthermore, password expiration can be “a source of frustration to users, who are often required to create and remember new passwords every few months for dozens of accounts, and thus tend to choose weak

passwords and use the same few passwords for many accounts,” also according to NIST observations [358]. Password policies exist to minimize the risk of user accounts being compromised, but password expiration tends to increase the risk. Password expiration is a defense against brute force attacks but not dictionary attacks [375]. Since other methods exist to protect against brute force attacks, password expiration policies should be avoided.

Another aspect to consider for protecting user accounts is the method by which passwords are stored and credentials are verified. Storing passwords unencrypted in a database creates risk because potential attackers can easily gain access to any account by compromising the database. To limit this risk, websites often utilize a hash function, which is a one-way algorithm designed to take variable-size input (such as a password) and generate a fixed-size number called a hash [14]. Hash functions are used because the hashes can be easily recomputed, but reversing a strong hash function to obtain the original input is computationally intractable. Credentials are verified by comparing the hash computed from the submitted login form and the hash stored in the database.

Hash functions, like other security measures, have their pitfalls. Since hash functions convert variable-size input into a fixed-size output, multiple input texts can theoretically be used to generate the same hash. This occurrence is called a collision. Most strong hashing functions are designed to minimize their risk [14]. For example, researchers at the Shanghai Jiaotong University discovered that collisions for the popular hashing function MD5 can be found using an IBM P690 machine in about an hour [402]. This resulted in MD5 being considered insecure. Ultimately, the risk of collisions is minimized by choosing a stronger hash function with a larger output hash.

Another issue with hash functions is the use of rainbow tables. Rainbow

tables are large, precomputed lookup tables of hashes to cleartext inputs which can be used to find passwords based on the cryptographic hash [297]. Since the security of a hash function comes from the computational expense of reversing a hash, a precomputed table of common passwords and their hashes saves the attacker much time and resources. The most common defense against rainbow table attacks is the use of a cryptographic salt, a “secondary piece of information made of a string of characters which are appended to the plaintext” before hashing [441]. Salts make password hashes more resistant to rainbow tables because they have higher information entropy and are therefore less likely to exist in precomputed rainbow tables [441]. Furthermore, a unique salt per password means that an attacker cannot simply use one rainbow table to crack the entire database of compromised hashes but instead needs to build a new rainbow table for each salt. Kioon, Wang and Das of the University of Science and Technology in Beijing list numerous countermeasures to rainbow table attacks, the strongest of which is to “use a variably located calculated salt including information outside the database and application code” [441]. This means that an attacker must break into the database, the application server, and also obtain a third source of information to crack the passwords.

#### **9.4.2 Discussion**

Civil Good’s original design document does not specify a password policy other than requiring two passwords on user signup [252]. However, this policy actually reduces account security. The research in Section 9.4.1 shows that since password security comes from the level of information entropy and that password class restrictions actually lower the amount of entropy. By providing two passwords (which likely generate two different cryptographic hashes), the user would cut the time and space requirements needed to crack his or her account in half. The purpose of the two-

password specification in the original document is to protect user privacy by not collecting a password reset email. As an alternative, email addresses could be made optional, and any addresses provided would be encrypted with the answer to a secret account recovery question.

Civil Good should still encourage users to choose passwords that are sufficiently long and unique. The Civil Good design document also does not specify whether passwords should expire. Since research in Section 9.4.1 shows that password expiration does not reliably increase security and is a nuisance to users, it is not recommended for Civil Good. It is also recommended that Civil Good have a high maximum password character length but not impose restrictions such as character class requirements.

To defend against offline rainbow table attacks in case the database is compromised, passwords in the Civil Good database should be uniquely salted before being hashed and stored. For maximum security, the salt should be composed of randomized information in the database, the application code, and some third external piece of information. Passwords should never be stored in plaintext, and policy should be written such that no Civil Good employee ever asks users for their account passwords. To protect against online attacks, Civil Good should implement an increasing rate limit for failed password guesses using geometric doubling. However, it should be noted that the rate limit should not increase indefinitely since that would allow attackers to deny service by simply guessing wrong passwords. Furthermore, the rate limit should be based on IP address or guest session token so that an attempt to cause a denial of service in this manner would not affect the user's ability to log in.



### 9.4.3 Conclusion

Having a strong password policy and utilizing proven hashing functions (such as SHA256) to securely store and verify passwords are key to protecting user accounts on Civil Good. For users to feel confident that their accounts are safe, Civil Good should have a clear password policy that encourages the use of strong passwords but does not restrict the size of the password search space. Furthermore, Civil Good should only require one password since multiple passwords make user accounts less secure. Regarding password recovery, email addresses should be made optional (but not required) so that users who so desire can maintain their anonymity, and those who specifically want to provide an email address for password recovery reasons have the opportunity. Email addresses would then be encrypted like any other personally identifiable information. Passwords should be hashed with unique salts that incorporate data from both the application server and the database server.

## 10 Privacy

We researched the major privacy concerns that a website like Civil Good would have to face, including the minimal requirements for tracking users, browser fingerprinting, and third party risks. We then proposed recommendations for resolving those concerns. We also drafted a privacy policy based off our research.

### 10.1 Minimal Tracking Requirements

In order for the Civil Good website to run optimally, it may need to track certain types of information. However, since one of Civil Good’s primary concerns is anonymity, care has to be taken in order to ensure that no information gathered that could compromise the privacy of any Civil Good user. This section analyzes the information that needs to be tracked, the methods by which to track the information, and the potential risks associated with using these methods.

#### 10.1.1 Background

One reason for tracking information is session management. The need for this is due to the hypertext transfer protocol (HTTP), a communication protocol used to transmit data over the Internet between a Web user (also referred to as a client) and a server. When the client clicks on a link in a Web browser, the browser usually connects to a Web server identified by the Uniform Resource Locator (URL) embedded in the link. The client sends out a request message to which the server sends back a response message. These are respectively known as HTTP “GET” and “POST” messages. In basic HTTP, every time a client sends a GET message a new connection is made that acts as if it were the first message sent. Therefore, there is no sense of state. Due to HTTP not having a sense of state, certain features of

Web applications within websites are can be significantly hard to implement [442, 185, 320]. In order to alleviate this problem, there are several common methods to implement sessions that keep track of state.

One of these methods is for the server to keep track of the user's IP address which is contained within HTTP messages. However, this can be unreliable since IP addresses may not be unique to a user. This is due to proxies and Network Address Translators (NAT). They are used to forward HTTP messages from users and can cause multiple users under the same proxy or NAT to have the same IP address [361, 442, 185, 320].

The next method of session management is to keep a HTTP cookie containing a session ID in the Web browser of the client. A HTTP cookie is information provided by a server in a POST message that is stored in a client's browser. When the client makes subsequent HTTP requests the cookie is sent with the HTTP request in the header. In order to keep track of a session, a unique session ID can be stored within that cookie. The contents of the cookie can be encrypted, the cookie can be set to have a limited life span, and the cookie can only be accessed by the domains allowed by the server who set the cookie. This all minimizes the risk of an attacker hijacking the client's session and consequently helps to maintain the privacy of the client. However, the server requires permission from the client's Web browser to use cookies. This method will not work if cookies on the client's browser are disabled [361, 442, 185, 305, 320].

Another method of session management is Hypertext Markup Language 5 (HTML5) Web storage. HTML5 is a language used for displaying content on a Web page. HTML5 contains a way to access client-side storage using JavaScript, a commonly used scripting language on the Web. This storage can be used to contain temporary information for session management. However, using malicious

JavaScript code, an attacker might be able to view the information and hijack the session, thus compromising the client's privacy. This is less of a problem when cookies are concerned because cookies can be set to HTTP-only which prevents them from being hijacked via malicious JavaScript code [442, 423, 305, 320, 360, 208].

A fourth method for session management is URL rewriting. In this method every URL from the user is modified to contain a session ID for the user. A server can then identify a user session based on the URL the message was sent from. However, since the URL is easily visible to others, this session ID can be effortlessly picked up by third parties. They can then use this ID to potentially hijack the original user's session. In addition, since every URL has to be rewritten, it can be a tedious process to implement [361, 185, 340, 174, 198].

Another method for session management is hidden form fields. These are fields input into an HTTP message. The fields contain information that is not visible to the client. In order to keep track of session, a session ID can be stored within these fields. The advantage is that the session ID is never stored to disk which makes it significantly harder for an attacker to steal. The disadvantage is that a form submission has to be dynamically generated for every page view which limits the usefulness of this method. For example, if the client clicks on a normal hypertext link a form submission is not generated. But, hidden form fields would work for a specific series of operations such as checking out from an online store.

A sixth method of session management is HTTP authentication. HTTP natively supports an authentication mechanism. Whenever a Web browser makes an initial HTTP request on behalf of the user, it prompts the user to enter credentials and then temporarily stores them. In subsequent HTTP requests these credentials are included. The server on the receiving end can then use these credentials to identify the session. The problem with this method is the user interface. Each time

a new session is started the client must enter credentials in a pop-up window. Little can be done to change the format of the window [361, 340, 198, 205].

The last method of session management is the “window.name” property in Web browsers. This property is accessible via JavaScript and can be used to store information about the user. However, this information can be shared among different tabs within the Web browser. Websites opened in other tabs could look at the information stored to find other websites where the client has been, thus potentially creating privacy concerns [361, 442].

In addition to session management, websites can track user patterns within their website to gather valuable usability data. This usability data can be valuable to development of the website. The patterns observed are generally not associated with individual users and so user privacy cannot be compromised [361].

Usability testing is generally accomplished through click stream analysis. This can be done by using JavaScript code to keep track of when a user clicked on buttons or links within a website. By analyzing the sequence of buttons a user clicked on the website, valuable data can be gathered such as which features on the website are most commonly used or the typical paths people use to access those features [361].

Another use for tracking is Web analytics. This is mainly focused on general usage statistics of websites. This includes monthly traffic data, where users are referred from, what types of users are interested in the website, and the effectiveness of marketing campaigns. All this information can be used as metrics to measure the success of a website and could be valuable to the ongoing development of a website. For example, Web analytics could be used to study user lifecycles on the website and find bottlenecks in the user lifecycles (the point where users tend to stop using the website) [315]. That bottleneck could then be addressed for the purposes of

attracting new users and retaining old ones.

In order to keep track of these statistics, self-hosted software or tools such as Google Analytics can be used [361, 442]. Unfortunately, Web analytics tools and user tracking techniques cause some privacy concerns. If a user reaches the tracking site through a public social networking profile, the site could potentially access that social networking profile thereby compromising the user’s personally identifiable information [361]. Google Analytics, which only uses first party cookies that are less capable of tracking across websites, has the same issue due to a technique called “cookie handover.” Cookie handover is a process where user IDs are communicated between domains using JavaScript code [361].

Instead, it is possible to gather usage statistics by analyzing server logs [361, 315]. Known as “basic Web metrics,” these statistics include the number of unique visitors and the time-ordered series of Web pages that the user views [315]. However, in comparison to more advanced Web analytics, this data is inadequate and can sometimes lead to inaccurate conclusions due to the data’s ambiguous nature [315].

### **10.1.2 Discussion**

Civil Good can potentially track certain information about users for the benefit of its website. However, one of Civil Good’s primary goals is to keep its users anonymous. As a result, Civil Good has to be careful when tracking any information. Any information gathered as well as the methods used to gather that information have to pose minimal risk to a user’s privacy.

One thing Civil Good needs to consider is session management. Since users will log into a profile and have ongoing conversations, a sense of ongoing state will be needed and therefore a means of session management will be required. There are seven methods of accomplishing this as mentioned in Section 10.1.1.

The first of these methods was to use the IP address of the user contained within a HTTP message. However this was unreliable due to multiple people having the same IP address under a proxy or NAT. Due to this, we recommend this not to be used as a primary means to keep track of sessions on Civil Good. However, it could still serve as a secondary, backup method in case other methods fail.

The second of these methods was to store a session ID inside of a cookie. This served as a more reliable means to keep track of session and also minimized the risk of compromising user privacy. However, users could potentially have cookies disabled on their browser. Despite this, a cookie containing a session ID could still potentially serve as a means of session management for Civil Good, provided that there is a secondary manner of tracking session in case the user has cookies disabled.

The next method is to store a session ID in HTML5 Web storage. However, there was the risk that if a website used this method, an attacker could hijack the session with malicious JavaScript code. If the session was hijacked, the attacker would have access to the user's profile and any private information stored in the profile on Civil Good. Thus we recommend that HTML5 Web storage not be used as a primary means of session management.

The fourth method was rewriting the URL to contain a session ID. This made the session ID easily visible, making the session significantly vulnerable to being hijacked and consequently risking user privacy. In addition, it was a tedious process to implement. Due to these disadvantages, we recommend against using URL query strings as a method of session tracking.

The fifth method was to store the session ID in hidden form fields. The form submission required for this method had to be dynamically generated for every page view. Since users of Civil Good will generally be navigating within their session by clicking on links, Civil Good will not be able to use hidden form fields in order to

keep track of the user's session.

The second-to-last method was HTTP authentication. This method had the restriction where users had to enter their credentials inside a pop-up window in order to start the session. Since this limitation did not occur with cookies we suggest giving cookies a higher precedence than HTTP authentication. However, HTTP authentication could be used as a backup method of session management if cookies are disabled on the user's Web browser.

The last method of session management was to store the session ID using the "window.name" property in Web browsers. This also created privacy concerns since other third parties had potential access to the property. Due to this, we recommend that this not be used as the primary means of session management for Civil Good.

In summary, for session management, cookies were generally better than the other choices available. HTTP authentication and IP address both worked as backups in the case the user has cookies disabled on their browser. IP address had some reliability problems due to users having the same IP address under a proxy or NAT. Considering all of this, we recommend cookies to be the primary means of session management on Civil Good, HTTP authentication to be a backup in case cookies fail, and IP address to be a backup in case both cookies and HTTP authentication fail.

In addition to keeping track of sessions, Civil Good also has the option of tracking the usage patterns within its own website using click stream analysis as mentioned in Section 10.1.1. Civil Good would not have to associate any usage data with any specific user, therefore providing no risk to the privacy of any individual user. Since the usability data provided could benefit the development of the website, we recommend that Civil Good gather this data.

Civil Good could also use tracking for Web analytics as described in Sec-



tion 10.1.1. Data to be gathered included where users are referred from, what types of users are interested in the website, and the effectiveness of marketing campaigns. Ideally, this general statistical information would not be associated with any individual user. Precautions should be taken to ensure that individual user data is not wrongfully collected. Since this data could be of value to Civil Good, we recommend that it be gathered. Since there are some risks associated with using tracking techniques for Web analytics on individual users, Civil Good might consider gathering usage statistics strictly through the server logs. While this data is not as useful or as accurate as obtaining full Web analytics, it avoids privacy risks caused by third party Web analytics tools and can still provide a rough guideline for Civil Good's further development.

### **10.1.3 Conclusion**

In summary, we make three recommendations for which information Civil Good should track. First, Civil Good should track user sessions through cookies, HTTP authentication if cookies fail, and IP addresses if HTTP authentication fails. Secondly, Civil Good should gather Usability data through click stream analysis. Finally, Civil Good should gather Web analytics through server logs. All the information gathered could prove beneficial to the functionality or development of the Civil Good website.

## **10.2 Browser Fingerprinting**

Application security is not limited to the backend server. Since Civil Good aims to protect user privacy, it should avoid storing any information that could potentially uniquely identify a user. Modern Web browsers provide identifying characteristics in every page request, and it is important that Civil Good determine the extent to

which the information it collects can be used to track individual users apart from obvious information such as date of birth and real name.

### 10.2.1 Background

The Electronic Frontier Foundation (EFF) conducted a study to determine the extent to which modern browsers are subject to “device fingerprinting,” [111]. Device fingerprinting refers to the practice of using information that modern browsers commonly send in page requests to uniquely identify a particular browser. Examples of such information include User Agent, HTTP ACCEPT headers, whether cookies are enabled, screen resolution, time zone, browser plugins, plugin versions, MIME types, and system fonts [111]. This study concluded that although fingerprints are not particularly stable (since many of these pieces of information can change over time), “browsers reveal so much version and configuration information that they remain overwhelmingly trackable” [111]. According to an article on PCWorld, browser fingerprinting can identify the user “far more accurately than any cookie” [224]. According to fraud prevention solutions company 41st Parameter’s Web page on “Complex Device Recognition,” device recognition can involve over 100 different factors [3]. Researcher Keiji Takeda from Keio University in Japan expands upon this study by examining other methods of online device fingerprinting techniques such as recording physical device Media Access Control (MAC) addresses, Bluetooth signals, and email header information [395]. Takeda’s methodology combines information from several different online device fingerprinting techniques to produce a more reliable and stable system for identifying unique devices.

In an article on CNET, Elinor Mills reports that online banks often utilize device identification techniques. She reports that although “none of the information gathered during a log-in is personally identifiable,” the bank should not have to col-

lect such data on a regular basis [276]. Jennifer Granick of the EFF is quoted in this article claiming that “there is very little privacy protection in the U.S. for this type of information” [276]. Mills also proposes one hypothetical example of a privacy breach that is possible with browser fingerprinting—banks who collect such information can determine whether certain user accounts are roommates or otherwise share a Web browser [276].

Collection of data is not without reason. Some information such as IP addresses can be used for network intrusion detection [40]. Cookies, time zones, language and browser types can be used to help track down fraudulent user activity [276, 3]. Regardless of the reasons such data is collected, researchers agree that collection of such data is a concern for user privacy and that both website privacy policy makers and Web browser developers alike should take browser fingerprint information into consideration [111, 395, 107].

### **10.2.2 Discussion**

Civil Good’s privacy policy should clearly state that it does not collect any personally identifying information, including anything that when combined can produce a uniquely identifying browser or device fingerprint. If any constituent information is stored (such as IP addresses or User Agents, which are common in Web access logs), Civil Good should take measures to ensure that such information is difficult to associate with a specific user account. Should Civil Good allow any third-party code to run on its site (such as advertisements), it should take steps to ensure that no such information can be collected by the third party entities, which will be discussed in greater detail in Section 10.3.

Some of the information could be useful without potentially violating user privacy. IP addresses, for example, can be temporarily logged in case they are needed

to investigate a security breach or other attack. Storing User Agent information can help Civil Good Web developers identify the browsers that are most popular and therefore most deserving of attention. As long as this information is kept separate from other fingerprinting information, creating a unique fingerprint should be reasonably difficult.

### **10.2.3 Conclusion**

Civil Good should develop policies such that no information that can be used for browser fingerprinting is logged or stored together. However, Civil Good should log IP addresses for at least a few days in case they are needed for investigating an attack.

## **10.3 Third Party Privacy Risks**

Keeping users anonymous is one of the primary goals in the design of the Civil Good website. In order to keep the anonymity of users, Civil Good needs to be aware of what third parties can do in order to potentially compromise the privacy of the users on its website. This section examines the methods third parties use to track information of Web users and then makes recommendations for Civil Good that help maximize user privacy.

### **10.3.1 Background**

Most third parties track users for advertising purposes. Many of these third parties, mainly advertising companies, try to build up profiles of users on the Internet. Profiles can include information such as age, gender, or browsing patterns. Whenever a specific user visits one of an advertising company's websites, the advertising company can use the profile information to display ads that the user would be more

interested in. In order to build up these profiles, there are several tracking techniques that can be used [186, 290, 361, 320].

The first of these tracking techniques is cookies. Cookies are information stored on the user end that can be set by websites. A website that sets a cookie is then sent these cookies whenever the user visits the website again.

First-party cookies are cookies that belong to the website that the user is visiting—in this case, the cookies used by Civil Good to maintain the authentication information. As described in Section 10.1, these cookies are benign in that they provide a service to the user (for example, ensuring that the user does not have to log back into Civil Good to perform an action every single time), and are not used to compile any sort of historical record of the user. It is important to note that even these cookies can pose a privacy risk, since they can be stolen by hackers to surreptitiously gain unauthorized access to a user account [63, 242].

If a third party owns or displays ads on multiple websites, cookies can also be configured to be sent to any of these websites when the user revisits them. These are known as third-party cookies, since they do not belong to the website the user is visiting directly [361]. Since third parties can display advertisements on multiple different websites, storing a cookie with the same advertisement server domain name for more than one website allows the third party to eventually build up a history of some of the websites the user has visited. Moreover, if any websites give the advertising company more information such as email, age, or name, the advertising companies can add this to their profile of the user as well [186, 361, 442, 320]. Thus, while cookies cannot carry viruses (that could steal private information), or install malware (that could read files on the browser's host computer), third party cookies can be used to compile a minimal historical record of the user's browsing habits at best and a comprehensive user dossier at worst [442, 361].

Zombie cookies are a special type of cookies that pose significant privacy concerns since the usual remedy of deleting all cookies does not affect them. A zombie cookie is a cookie file capable of recreating itself after the user carries out a cookie deletion action [162]. This is accomplished by storing a backup of the cookie in storage locations that are not easily accessible by the user, such as the local storage available for Adobe Flash content, and recreating the cookie from these backups when required [379, 162]. The privacy risk is born because of the fact that “the removal of Zombie cookies is tedious and requires substantial effort” [361].

Another technique advertising companies can use to add information to a user’s profile is to look at the HTTP referer header. The HTTP referer header tells what link the user clicked on to get to a website. Advertising companies can use this to tell where a user came from when the user visits any part of their website. They can add this information to the profiles they have built up [308, 361, 442, 320]. However, if the website the user was sent from uses HTTPS, this field will be left blank [308].

A third technique advertising companies can use to track users is Web bugs. These are objects embedded in HTML that are invisible to the user such as a one pixel by one pixel image. When a user loads a website page containing these objects, a request is made to the advertising company’s servers to load these objects. From this, the advertising company knows when the user has accessed the page containing the Web bug and can add this information to their profile on the user [186, 290, 361, 320].

Another technique that can help in building the advertising company’s profile on the user is browser fingerprinting. This technique, as described in detail in Section 10.2, can help to uniquely identify users by gathering information about their browser. Advertising companies can potentially associate this fingerprint with

a user's profile. Thus, advertising companies can use a user's browser fingerprint to figure out which of the advertising company's websites the user is visiting [186, 361, 442, 320].

The last technique advertising companies can use to add information to user profiles is through IP addresses. Every time a user visits a website, the user's IP address is given to the website. If the user is not under a proxy, geographic location can be inferred from IP address. Advertising companies can add the geographic location and user IP address to the profile on a user if the user visits one of the advertising company's websites [186, 442].

### **10.3.2 Discussion**

One of Civil Good's primary goals is anonymity. In order to achieve this, Civil Good needs to be careful in its design to prevent giving away private user information to third parties. Civil Good can take several measures against the tracking techniques mentioned in Section 10.3.1 in order to ensure privacy.

The first of the tracking techniques mentioned was cookies. Advertising companies could use cookies in order to tell whether a user has been to one of their websites. However, the only information they could gather from this technique is whether or not the user has been to Civil Good. As long as Civil Good does not reveal further information to the third party about the user, then this further information cannot be gained. Allowing advertisers to display certain ads to a targeted demographic is of great value to the advertisers, and thus they would be willing to pay more. Section 12.4 will describe this in greater detail. However, if the advertisers were to know the users came from Civil Good, they could then associate the demographic they targeted with the user. In this case, in order to prevent privacy risks, Civil Good would have to disallow cookies from the advertiser. This

way, the advertisers would not know the user came from Civil Good. Consequently, Civil Good has no way to prove to the advertisers that the advertisement caused users to visit the advertiser's website.

To mitigate the privacy risks caused by zombie cookies, it is recommended that the architecture for Civil Good's website prevent cookie access to any extra-neous storage above the space normally provided by a browser to store cookies. The user can also be warned about the risks and asked to configure their browser privacy settings to block third-party tracking cookies entirely in order to prevent zombie cookies from being set on the browser.

The second tracking technique was Web bugs. Web bugs would generally provide the same information as cookies with regards to informing a website whether or not a user has been to Civil Good. Thus, Civil Good has the similar options for dealing with Web bugs as it does for cookies. Civil Good can allow Web bugs from third parties and not provide any more information. Alternatively, Civil Good can disallow Web bugs and advertisers but would have no way to prove the success of the advertisements to advertisers.

The next tracking technique was the HTTP referer header. Civil Good should use HTTPS for all requests made to its servers. Therefore, the referer header will be left blank and this should not be a concern for Civil Good.

The fourth tracking technique was browser fingerprinting. In order to avoid this risk, Civil Good should limit the code a third party can execute from the Civil Good website. This limits the ability of a third party to recognize the browser fingerprint of a user.

The last tracking technique was IP address. This would unavoidably be sent to the advertisers whenever a user follows an advertisement. From this, the advertisers could gain the IP address itself and the geographic location of the user



if the user is not under a proxy. If Civil Good were to advertise, it should mention in its privacy policy that if a user clicks on an advertisement, they do so at their own risk

### **10.3.3 Conclusion**

We make three recommendations for Civil Good. First, Civil Good should either (1) allow advertisers to use cookies and Web bugs but not reveal any more information about users to the advertisers or (2) allow advertisers to target a specific demographic but disallow cookies and Web bugs at the disadvantage of not being able to prove to advertisers that the advertisements are effective. Next, Civil Good should limit the code a third party can execute to prevent browser fingerprinting. Lastly, if Civil Good use advertisements, it should add to its privacy policy that it is not responsible for the user's IP address and therefore geographic location being determined by a third party if the user clicks on an advertisement.

## **10.4 Privacy Policy**

Civil Good will contain a wealth of information about a specific user's views on a wide range of subjects, as well as play host to a fair amount of demographic information about that user. It is thus of paramount importance to clearly communicate to the user Civil Good's policy about the privacy of the user's information.

### **10.4.1 Background**

The Internet relies on a self-regulatory approach to privacy. With the wide range of choices available to Internet users, it is assumed that users will gravitate towards websites that maintain "acceptable privacy policies" [263]. It is also assumed that websites will uphold their own privacy policies because of the potential for repercus-

sions from governmental actors should the companies choose to do otherwise [263]. The legal issues surrounding website privacy policies are discussed in Section 5.

Formulating a privacy policy helps to increase the trustworthiness of websites [59, 20]. In 2004, “an estimated 77% of websites” posted a privacy policy [201]. With the recent explosion of articles in mass media about the data collection campaign by governmental agencies, it has become increasingly important to reinforce user privacy.

A privacy policy should address issues “that users care about” [201]. There is a discrepancy between what typical privacy policies address and what are the most pressing concerns for users [110]. Some privacy policies have been criticized for not providing any real value to the user [277].

Privacy policies that rely heavily on legalese typically have a lower number of users actually reading through the entire policy. Even if the user does read the privacy policy, there is no guarantee that the policy will be understood by the user. Quite a few studies have concluded that users must possess the reading level of at least a college student in order to be able to understand privacy policies [263, 277, 161]. If fewer users are cognizant about the website’s privacy policy, it is detrimental both to those users’ experience and to the general culture that would develop on the website as a whole, since being unaware of the ground rules regarding the log of the conversation would not be conducive to sharing deeply-held personal opinions. This would, in turn, reduce the level of user-perceived value from Civil Good. The effects of a privacy policy—namely, to “promote consumer choice and reduce the risks of disclosing personal information online” [277]—will be lost if the privacy policy is ignored by users.

To ensure that the privacy policy is read by users, the policy should also be presented “in a format that they (the users) comprehend and subsequently

use” [277].

### 10.4.2 Discussion

Any privacy policy developed for Civil Good should have two inviolable basic tenets: the policy must address concerns that would be relevant to a user, and the policy must be strictly enforced. If Civil Good were to address the most pressing user concerns with a privacy policy in a transparent manner, the website’s reputation would increase. In addition, when users feel that their privacy is protected and respected by the website, it can help to drive up user retention rates, since users typically use websites that they trust more frequently [138]. It can also contribute to an overall positive user experience.

It is essential that the privacy policy be stated without extraneous legal language. The format of the policy should also be one that presents the relevant information in an easy to read manner. This encourages users to read and understand the policy. If users are informed about the privacy policy from the very outset, they are more likely to feel safer when sharing their information.

With these guidelines in mind, we developed a privacy policy for the Civil Good website. The privacy policy is listed in Appendix E. We developed a policy containing a list of bullets so that the policy would be easily readable. The privacy policy was written with the aim of avoiding extraneous legalese, and was reviewed by Stuart R. Malis, Esq. as an appropriate initial draft policy. The policy will have to be reviewed and finalized by legal counsel for Civil Good before the site goes live, and will have to be amended and updated by Civil Good legal counsel over time, as required by legal changes or Civil Good policy changes.

We also recommend following the generally observed practice that a hyperlink to the privacy policy be placed in the footer section of the website. Since the footer

is displayed on each page of the website, this will provide users with easy access to the privacy policy at any point during their interaction with the website.

### **10.4.3 Conclusion**

We have provided a proposed privacy policy in Appendix E. Incorporating the privacy policy into the website will help generate trust, keep users informed about what they can expect from the website in terms of privacy, and will improve user retention.

## 11 Community - Developers and Users

We examined methods of attracting developers to a project like Civil Good with regard to what motivates software developers to work on nonprofit projects. In addition, we looked at the various ways of attracting and retaining users, which is key for a website to sustain itself over time.

### 11.1 Attracting Developers

Civil Good is intended to be a not-for-profit service. By extension, Civil Good would greatly benefit from the reduction of costs in its development. While it is entirely possible to hire full time employees for the purpose of development, there is a less costly alternative.

Open source software is computer programming code that is available publicly for usage, modification, and sharing by anyone without the requirement of any monetary exchange [341, 188]. Within the open source community, there are individuals willing to contribute to projects for a variety of different reasons ranging from their personal passion for the work to the recognition to be gained from successful projects.

#### 11.1.1 Background

In both open source projects and proprietary projects, developer incentive can be divided into two categories: immediate payoff and delayed payoff [235]. Immediate payoff is almost always in the form of monetary compensation. In comparison to proprietary projects, monetary compensation is something that open source projects lack. Instead, open source projects tend to appeal to developers with some form of delayed payoff.

Peer recognition is one such form of delayed payoff [235, 82]. Developers with valid contributions to successful open source projects build a strong reputation among other developers. As opposed to proprietary software where the actual contributions of an individual cannot be directly examined, open source software offers a chance to see exactly what type of contribution was made and how effective the solution was. High quality contributions are more visible in this way and are a significant factor to improving a developer's reputation [221]. Any computer code available under an open source license also allows "modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software" [188]. This allows a developer to present their work to the wider community as a foundation to build upon, thereby further fostering relationships and adding to the developer's reputation.

Of course, recognition can only come with a project that has an audience. The main target audience in this scenario is other developers [235]. Generally, the types of projects that have a developer audience are projects that developers would likely make use of such as operating systems or programming languages [235]. Linux, a leading open source operating system, and the Apache Web server, a leading open source Web server, are two popular examples of these types of projects [135, 235].

Other projects that garner a developer audience are projects that developers perceive as having strong potential. If a developer believes that many other developers will be interested in working on a project, then that developer is more likely to contribute to that project [235]. However, a project has to be well-established enough to reach that degree of recognition. There must be strong evidence, often in the form of an initial codebase, that the project's concept is viable and the project has the capacity for extensive growth [235].

For example, the Apache Web server originally came about due to a growing

need in the Webmaster community. With slowed development on the leading Web server software at the time, Webmasters started writing their own extensions and fixes to the original server software. These additions were eventually consolidated into the Apache Web server [280, 135]. The viability of the Apache project was proven with its initial codebase: Rob McCool's HTTP daemon (httpd), the most widely used Web server before the Apache Web server [135]. The project's growth potential was clear to the developers at the time since there were already several contributions by individual Webmasters leading up to its initial release.

In addition to the payoff developers receive for contributing to an open source project, it is also important to consider the costs developers incur for an open source project. Potential developers will only contribute if the benefits of contribution outweigh the costs [176]. In most cases, the immediate cost of contribution is time [235, 176]. Open source development has three advantages over closed source development that help to reduce the cost: personal benefits, career benefits, and the "alumni effect," which is the reduced learning curve resulting from having a project openly available to academic communities [235].

Developers who have full discretion over what projects they work on (which is the case for open source development) are likely to gain private benefits for their work. Since developers are attracted to projects that they use privately, then they have an inherent benefit in improving that project.

Human capital theory explains open source participation as a way for developers "to gain marketable technical skills" [176]. This is another form of delayed payoff. Developers can improve their developing ability either by acquiring new development skills or improving old ones. In this way, developers can use open source projects as an outlet for building their resumé in the hopes that the projects will improve their work opportunities in the future.

The “alumni effect” reduces the learning cost of contributing to an existing project. Since the codebase is free to be used for learning purposes, developers coming into an open source project may already be familiar with the codebase and will not have to spend time learning how to work with it. There is a significantly reduced time cost compared to any other project that does not benefit from the alumni effect.

Another reason that can spur developers to contribute to open source projects is the feeling of contributing to the greater good [176]. This is an example of the philanthropic mindset, where working on a project that achieves a societal change is a manner of giving back to one’s community. Here, that community is an online community that stretches passed geopolitical boundaries.

A final reason that motivates developers to contribute to open source projects is that they “enjoy the work of programming itself” [221].

### **11.1.2 Discussion**

If Civil Good chooses to attract developers as an open source project, a barebones (but functional) codebase is recommended to prove its viability and, at the same time, allow new developers a clear cut set of development tasks. Whether the initial codebase is built using a student workforce or using paid developers, having a codebase as described is essential to open source success.

On the topic of development tasks, it is important that the codebase be compartmentalized well enough that individual tasks are well-defined and mutually exclusive from other tasks [235]. The remaining work should also specifically include non-menial (i.e. challenging) tasks due to the powerful delayed incentives these tasks might offer.

Incentivizing through peer recognition and employer recognition, as discussed



in Section 11.1.1, is significantly more effective when the development tasks have some weight to them [235]. Developers will not be able to demonstrate their ability on menial user-focused tasks such as documentation or user interfaces [235]. In addition, developers gain more valuable technical skills working on development tasks that do not have a readily apparent solution or that require additional learning on the part of the developer. As an aside, it is important to note that menial tasks can still be outsourced through the open source model since there are other incentives to perform these tasks [221].

Peer and employer recognition is an especially strong incentive early in the project's development. With less actual code in the codebase, high quality contributions have increased visibility compared to a project that has already thrived for several years. Items such as the addition of full features have an impact that is both quantitatively and qualitatively greater with respect to the growing codebase. Early development is a period when developers have the opportunity to establish themselves as key figures in the success of the project. By extension, the importance of the project's presentation at this time is critical.

Some possible tasks that could be left for the open source community are the details of the member searching process, the data mining algorithms for statistical queries, and the implementation of the sorting of archived conversations.

Pertaining to the construction of an initial codebase, the use of open source software is encouraged due to the possible benefits of the alumni effect discussed in Section 11.1.1. Software products that are highly used as learning tools are especially encouraged. The more familiar potential developers are with Civil Good's initial codebase, the less costly it will be for them to begin contributing and the more likely they will be to contribute. If possible, Civil Good should use this approach in order to take full advantage of the alumni effect.

### **11.1.3 Conclusion**

Open source development is a phenomenon in the software industry that can be just as practical as traditional development. Civil Good’s social and technical qualities make it a decent candidate for this style of development. If Civil Good is engineered correctly at its creation, open source development could be an effective source of code production. As an ending side note, while this examination looks specifically at principles of open source development, many of the same principles can be used to market private development.

## **11.2 User Trends and Retention**

In order for any website to prosper commercially or socially, it needs to be able to generate content that attracts and retains users. In the case of Civil Good, much of the work in attracting users is covered with its novel approach at online discussion. The more difficult challenge facing a new and growing website like Civil Good is creating and maintaining a loyal user base.

### **11.2.1 Background**

To establish Civil Good’s potential target audiences, it is helpful to consider the politically active side of the population. In a 2002 article, Scheufele et al. discuss the correlation between socioeconomic status and level of political participation noting that citizens of higher status are more likely and more willing to participate [362]. Scheufele suggests that people with higher education, people who are more involved with their communities, and people who make use of public media are most prone to political participation. Of note is the fact that media consumption through newspapers has a greater correlation with political participation than media consumption

through television [362].

Anne Schneider and Helen Ingram also discuss groups that are not politically active in their 1993 political science article. They suggest that groups with the lowest rate of participation are the same groups that have the most to gain from political action [359]. Specifically, they refer to the poor and to the minority groups. Instead, it is more often the “advantaged” groups that have a high rate of participation. Schneider and Ingram name advantaged groups as those including the elderly, businesses, veterans, and scientists [359]. This falls in line with Scheufele’s observations on socioeconomic status and political participation.

A study conducted by Panciera et al. on a website called Cyclopath examined user participation [303]. Cyclopath is an open content system built for providing and rating cycling paths for users in Minnesota. Panciera et al. conducted the study on the activity and usage habits of Cyclopath’s users during a critical point in Cyclopath’s development (i.e. 16 months after its launch) [303]. Specifically, they examined the effects of automatic login on user activity, the “life cycle” of Cyclopath users, and the rate of retention for different groups of users.

By default, Cyclopath has non-expiring login sessions; that is, users’ login information is stored indefinitely such that the user never has to re-enter it. This function was turned off over a two week period for Panciera’s study. Usually, 47% of users were logged in while performing site browsing activities and 90% of users were logged in while performing content revision activities. During the two week testing period, these numbers dropped to 24% and 62% respectively [303]. User activity while logged in was found to be significantly reduced when users were forced to log in manually.

On user life cycles, Panciera et al. examined the time elapsed from when a user begins browsing Cyclopath to when they perform their first anonymous edit,

and from then to when they decide to register. They found that few users participated in anonymous editing (only 18 out of 286) as most users would rather register before editing. This suggests that “anonymous user” functionality does little to entice users into registering. The study also found that 45% of users began contributing content within the first day of browsing the site while 44% of users waited a week before contributing [303]. Users were just as likely to register upon discovery of the site as they were to put off the registration and remain passive users.

On user retention, Panciera et al. found that, after 100 days, approximately 50% of contributing users were retained while non-contributing users were only retained 30% of the time [303]. It appears that users who contribute to a site are more likely to return in the future. Note that Panciera loosely defines a retained user as a user who revisits the site at least once some  $n$  (in this case, 100) number of days after registration.

Jenny Preece published an article on sociability and usability in online communities in 2001. She identified three essential components to a socially successful online community: a group of users (people), an interest shared by the group (purpose), and a set of protocols accepted by the group (policies) [323]. In particular, Preece discussed how policies created a sense of history among the community and how they can shape the community’s accepted social norms. On usability, she mentions the potential deterrent effects of user registration. If the registration process is rigorous, it will prevent users from participating in “hit-and-run flaming” [323]. At the same time, such a registration process can also drive away legitimate users from the community.

Bonnardel et al. conducted a study on the effects of color on website appeal in 2011 [42]. They constructed 23 different versions of the same homepage in 23 different hues and had study participants rate each homepage on a seven point scale.

Participant ratings concluded that the colors blue and orange were most popular amongst all groups. Following up on these findings, Bonnardel examined the effects of the colors blue, orange, and grey on the website users' information processing. They determined that users navigating the grey site spent less time on the site and recalled less information than users navigating the blue or orange sites. Users navigating the orange site also recalled the greatest amount of information compared to the other two sites. Note that the findings of Bonnardel et al. contradicted conclusions from a 2009 study that stated that website users had more of a preference for cool colors such as blue compared to warm colors like orange [84]. The color blue was still a popular choice, but the study by Bonnardel et al. was the first to put orange in the same category.

### **11.2.2 Discussion**

One of the main goals of Civil Good should be getting people to participate in their first one-on-one discussion. As was discussed in Section 11.2.1, users who make a contribution to a website are more likely to return to it in the future [303]. Based on this correlation, it is recommended that among the predefined conversation topics should be some visible set of “entry-level” (more accessible) topics. This would allow for more of the general public (and perhaps the less politically-inclined) to quickly get started using Civil Good.

On the topic of Civil Good's entry barrier, there should be some consideration set aside for the difficulty of registration. Jenny Preece's article discussed in Section 11.2.1 warns that users may be turned away by a registration process that is too difficult [323]. Civil Good should attempt to find a good balance that discourages throwaway accounts, but is still simple enough that potential new users are not alienated [102].

Preece also suggests that growing communities would benefit socially from some behavioral protocol or set of rules. Rules give a community a sense of presence even when that presence is not reflected by the size of the community. Preece also emphasizes garnering a sense of history within online communities that helps establish the community's social norms. Since Civil Good is not a system that supports forum-like discussion, building a sense of history might be difficult. However, updates presented to the Civil Good community (on behalf of its administrators) in a form similar to developer blogs might alleviate this problem.

Civil Good should also consider implementing a form of account memorization (e.g. a "Remember Me" checkbox on the login page). According to the Cyclopath study discussed in Section 11.2.1, removing the extra step of constantly re-entering account information between sessions significantly improves user activity. Increased activity in turn generates a stronger sense of commitment from the users and increases user retention.

Bonnardel's study stresses the importance of a website's aesthetic design [42]. Not only are visually attractive websites more pleasing to users, they also increase average website usage time and overall quality of use as shown by the second study conducted by Bonnardel et al. Something as mundane as the color of a specific banner on the homepage can have a significant impact on user perception of the site. Specifically, the color blue seems to have universal appeal among website designers and users [42, 84]. It is interesting to note that popular social media sites Facebook and Twitter predominantly use blue in their color schemes. Civil Good's designers should be aware of the consequences of their aesthetic design decisions.

### **11.2.3 Conclusion**

While there are several actions that Civil Good’s designers can take to guide the service to a sufficient user base, it ultimately falls on the users themselves to be able to create the type of content that allows online communities to thrive. That said, the responsibilities of Civil Good’s creators span from basic website design to the policing of the site’s users to promoting active participation within the community. These duties, if performed adequately, will hugely improve the website’s chances of growth and success.

## 12 Business Model

We investigated the costs of developing the website, operating it, and marketing it using various techniques. We developed a business model that could be used to meet those costs, and explored the advantages and disadvantages of becoming a 501(c)(3) nonprofit organization.

### 12.1 Operational Costs

Civil Good is a Web-based service, and most Web-based services can have several different types of operational costs associated with them. These costs include Web hosting and domain name. As discussed in Section 7.1, Civil Good should utilize VPS for its Web hosting. This section will expand upon that and discuss the details of the potential costs of Web hosting and domain name for Civil Good.

#### 12.1.1 Background

Civil Good’s VPS Web hosting is available through a variety of hosting providers. Each company generally offers different monthly plans based on the amount of bandwidth (BW), disk space, and RAM needed. As mentioned in Section 7.1, bandwidth is the size of the amount of allowable data transferred through the server per month by users of the website and disk space is the amount of space needed to store all information from the site. Lastly, RAM is the amount of temporary memory needed in order for programs to run [72, 396, 48]. CompareVPS provides a large list of these VPS plans [71].

Domain names are also offered by several different companies on online sites. Domparison gives a list of these companies for each type of name (such as ‘.org’, ‘.com’, ‘.net’) as well as the yearly prices each company offers [103].



### 12.1.2 Discussion

Before Civil Good can choose a plan for VPS, it needs to consider its bandwidth, storage space, and RAM needs. The first item, bandwidth, can be roughly extrapolated from observations of average Internet traffic per month on political websites similar to Civil Good. Column two of Table 3 gives monthly traffic data in terms of average total page views per month for some of these websites [409]. The average number of page views per month for each of the websites given in the table was about 21,000. According to Google, the average page size of a text-based website is about 320 kilobytes (KB) [98]. Thus, the bandwidth requirement is 320 KB times 21,000 which is approximately equal to 6.72 gigabytes (GB) per month. Most providers listed on CompareVPS provide at least 100 GB of bandwidth so this requirement should not be a concern.

Table 3: Traffic Data

Site URL	Pageviews	Unique Pageviews
thepoliticalforums.com	78,346.5	6,006.9
thepoliticsforums.com	51,825.1	7,416.8
liberalvaluesblog.com	18,709.2	14,672.7
usmessageboard.com	17,470.2	534.5
wyblog.us	12,077.1	9,898.9
thedemocraticdaily.com	6,131.7	4,675.2
opinion-forum.com	3,888.3	3,228.5
beyond-the-political-spectrum.blogspot.com	648	493
letpoliticstalk.com	276.7	189.2
Average	21,041.2	5,234.98

Pageviews are average per month.

The second item, disk space, is determined mainly by the size of the database for Civil Good. As described in Section 7.2, the estimated size per registered user was about a couple hundred KB. At most this would be around 1 megabyte (MB). Column two of Table 3 gives the average monthly unique pageviews for websites

similar to Civil Good [409]. A unique page view is only counted once per person's IP address. The average of these values, 5234.98, can give a rough estimate number of unique visitors expected per month. The number of these unique visitors that would sign up is given by conversion rate. The conversion rate for media site which Civil Good can be considered a subset of is 10% [51]. Considering all these conditions, the memory needed to support Civil Good for one year is 12 months times 5234.98 unique visitors per month times 1 MB per user times 0.10. This equates to about 6.2 GB. Thus to support Civil Good for a significant amount of time such as 8-10 years, we recommend about 50 - 60 GB of disk space to be included in the plan purchased.

For RAM, CompareVPS recommends the minimum amount to be 256 MB for a website [72]. We recommend this be used a minimal baseline for the RAM needed for Civil Good.

Table 4 gives a summary of various plans as found on CompareVPS that offer the approximate minimum bandwidth, disk space, and RAM requirements needed for Civil Good [71]. As seen in Table 4, the average price for each of the VPS plans is \$34.47 and should be around what Civil Good should expect to pay for a VPS plan.

Domain name costs are much simpler to estimate. Civil Good has a '.org' domain name. The average price, calculated from the list of '.org' domain name yearly prices on the website is \$11.63 [103]. This should be a good estimate of what Civil Good would pay per month for a domain name.

### **12.1.3 Conclusion**

The operational costs for running Civil Good were Web hosting and domain name. From the discussion, the respective costs for Web hosting and domain name are

Table 4: Web Hosting Prices

Hosting Company	RAM (MB)	Disk Space (GB)	Bandwidth (GB)	Monthly Cost (\$)
VPSFX.com	256	50	1,000	3.25
Azza VPS	1,024	65	20,000	4.73
Cloud Shards	1,024	50	900	7.44
VirtualBlock	1,024	50	1,000	9.45
TransIP	1,024	50	1,000	20.00
CheapVPS	768	50	1,200	30.79
VPSNext	384	50	1,000	39.34
RootBSD	1,024	60	300	47.23
Host Color	768	50	350	55.09
HostV	256	50	1,000	55.12
HostForWeb	1,024	60	2,000	65.63
easyCGI	1,024	50	200	75.58
Average Cost:				34.47

about \$34.47 per month and \$11.63 per year.

## 12.2 501(c)(3) Status

Through an application process, the government grants 501(c)(3) tax exempt status to certain nonprofit organizations that operate for the general interest of the public [366]. Civil Good operates for improving social interaction on the Internet for the benefit of the public. Because of this, it may have the potential to become a 501(c)(3) organization. This section will provide an overview of the potential process of Civil Good becoming a 501(c)(3) organization and analyze the advantages and disadvantages of doing so.

### 12.2.1 Background

Information about becoming a 501(c)(3) nonprofit organization is given in Internal Revenue Service (IRS) documents: Publication 4420, Publication 557, and Form 1023. First, an organization must meet a specific set of requirements to attain 501(c)(3) status. Next, the organization must submit an application to the IRS. If

the application is approved, the organization must then meet certain responsibilities in order to maintain 501(c)(3) status [366, 367, 365].

The main requirement the organization has to meet is that its purpose is to operate solely for the good of the public. Nonprofit organizations that meet this requirements usually fall under the categories of charitable, educational, or religious organizations [366, 367]. Next, the organization must have a set of organized documentation that limits the operations and assets of the nonprofit solely for the use of furthering its purpose [366, 367]. Third, the organization may not partake in activities that further a politician's political campaign, or that benefit its founders or private shareholders [366, 367]. Lastly, the nonprofit must be organized as a corporation, unincorporated association, or trust [366, 367]. A corporation is a separate legal entity recognized by the government, independent from its owners. Members of a corporation do not hold liability for actions of the corporation [130, 281]. An unincorporated association is a group of people who work for a common purpose [130, 281]. A trust is the arrangement where one person is given property by another where the property is to be used for the benefit of a class of persons or the general public [130, 281].

After a nonprofit organization has met the requirements, it must then apply for an Employer Identification Number (EIN) through Form SS-4 [366, 367]. An EIN is "required regardless of whether the organization has employees" [366]. After receiving a number, the organization then starts the application process for 501(c)(3) status by filling out Form 1023 and paying a user fee to submit the form. The user fee is either \$400 if the organization's income is less than \$10,000, or \$800 if the organization's income is greater than \$10,000 [366, 367, 365]. Additionally, Form 2848 and Form 8821 are required if someone other than the organization's principal officer or director will represent the organization in all matters related to

the application [366, 367].

If an organization's application is approved, the organization must meet certain responsibilities in order to maintain its status. First, the organization has to file an annual information return. It must disclose the contents of this information return and Form 1023 to the public [366, 367]. Additionally, the organization must keep records of all financial and non-financial activities [366, 367]. Lastly, the organization has to send a written acknowledgement to donors with contributions of \$250 or more and a written disclosure to donors with contributions in excess of \$75 [366, 367].

The benefit an organization receives from the government from 501(c)(3) status is exemption from federal income taxes [366, 4]. Secondly, charitable contributions from donors are eligible to be tax-deductible. Because of this, other corporations or individuals may be more likely provide support [366, 4]. Lastly, the organization may receive additional exemptions from state or other local taxes [366].

In addition to government benefits, businesses may also offer free or discounted services to 501(c)(3) nonprofit organizations. One such example is Google Adwords. The program offers up to \$10,000 per month in free advertising to 501(c)(3) organizations [154]. Another benefit is with Web hosting. Some companies offer free or discounted Web host plans to 501(c)(3) organizations [163, 70, 105]. Lastly, some law firms offer cheaper services for 501(c)(3) organizations. Their services are to help the organizations become established and maintain their 501(c)(3) status [78, 53, 5].

### **12.2.2 Discussion**

Civil Good has a strong chance of meeting the requirements to be a nonprofit. Its purpose, operating for improving social interaction on the Internet for the benefit

of the public, falls under the category of education which likely makes it eligible. In fact, one of the common types of educational organizations listed by the IRS is described as “organizations that conduct public discussion groups, forums, panels, lectures, or similar programs” [366].

A choice Civil Good would have to make before becoming a 501(c)(3) non-profit is whether to be established as a corporation, unincorporated association, or trust as described in Section 12.2.1. The advantage to the latter two is that they are less organized and thus simpler to form. However, this advantage is partially mitigated by the fact that Civil Good would have to alter its organization anyways to achieve 501(c)(3) status, as mentioned in Section 12.2.1. The advantage to being a corporation is the protection of its members from liability. Additionally, the stronger organization of a corporation could potentially help in filling out the organization documents for 501(c)(3) status. Taking into consideration the advantages of each option, it is suggested that Civil Good become a corporation if it were to obtain 501(c)(3) status.

After becoming a corporation and receiving 501(c)(3) status, Civil Good will receive governmental and business benefits as mentioned in Section 12.2.1. However, along with these advantages come several disadvantages that Civil Good would have to take into consideration.

One of these disadvantages is that, as described in Section 12.2.1, Civil Good would have to be a corporation in the United States to be eligible for 501(c)(3) status. This fact subjects Civil Good to the laws of the United States Government. This could lead to potential legal issues as described in Section 5.

Additionally, as described in the background, there are legal paperwork requirements throughout the application process, such as the organizing documents, Form 1023, and the annual information return. This requires significant time and

effort to manage. However, the help offered by some law firms may help to mitigate this disadvantage.

Another, smaller disadvantage is the user fee required to fill out Form 1023. This is a small fixed fee and can be paid off given enough time.

One final disadvantage is that Civil Good has to already be established before being considered for 501(c)(3). This means that Civil Good cannot depend on 501(c)(3) status in its initial phases of development. However, if Civil Good does become an established corporation, the potential benefits provided by 501(c)(3) status could still help ease the growth of the website.

An additional consequence Civil Good should consider is not being able to support a specific politician's campaign. Extra precautions would have to be taken in order to ensure that one politician is not favored by the website over another. Debate on the website by the users over which politicians are better should still be allowed as long as Civil Good does not endorse the debate in one particular direction. Although this complicates Civil Good's operations to some degree, it also provides extra incentive to avoid creating a politically polarized community (which is potentially damaging to the purpose of the site) as discussed in Section 6.1.

### **12.2.3 Conclusion**

Despite the issues associated with becoming a 501(c)(3) nonprofit organization, it is still beneficial to obtain 501(c)(3) status. 501(c)(3) status requires much legal work and an established base; however, it significantly helps to alleviate the costs to run Civil Good. Thus, in order to benefit from these advantages, it is suggested that Civil Good apply for 501(c)(3) status after it has been established.

## 12.3 Marketing Costs

Civil Good could not succeed without its potential community members learning about it. Although not a technical operational cost, marketing campaigns can easily become an important investment. The most common forms of online advertising are search engine, display network, and social media marketing. Costs vary among each type, and each method has its advantages and disadvantages. However, online discussion is not a high-volume, high-competition campaign, so overall costs are roughly the same among the three types. What then becomes important is choosing the methods that bring users who monetize better for the same cost.

Another manner of advertising that is prevalent on the Internet is email marketing. While it is different from the first three forms in that it is a more personal communication with a customer, it is a method of acquiring customers for relatively low cost.

### 12.3.1 Background and Discussion

Search Engine Marketing (SEM) is the practice of running online text advertisements in the results of popular Web search engines. Advertisers bid on certain keywords, and when users search these terms, the search engine displays relevant advertisements. Google AdWords (a major provider of online search engine advertising) has a tool to help estimate SEM advertising costs, which has been extensively used to roughly estimate the SEM advertising costs of a website like Civil Good.

Online debate is not widely seen to be a lucrative market, so there is very little competition. Table 5 shows that cost estimates are very low; the suggested bid from Google for online ranges from 0.24 to 0.56 per click. Average volume is less than 1,500 impressions per day and less than 50 clicks per day. From the Figure 1, cost growth is logarithmic and a rough estimation for total SEM costs anywhere less



Table 5: Results from Google Keyword Planner

Keywords	Suggested Bid	Searches	Competition
debate, online debate, online discussion	\$0.24	714,770/mon.	low
One-on-one debate	\$0.56	140/mon.	low

Suggested bids, average search per month, and competition levels of keywords related to online debate [156].

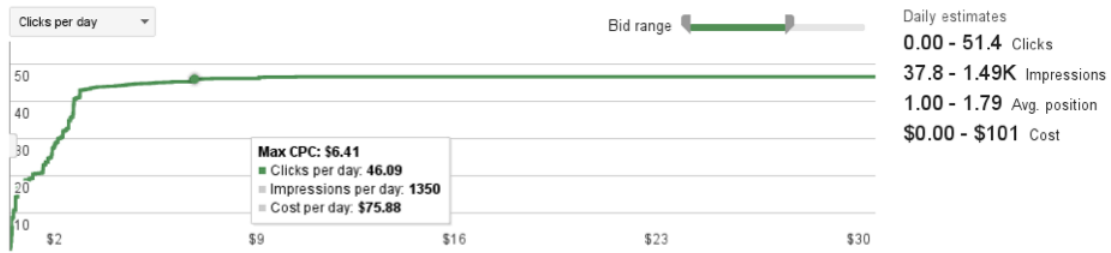
than \$100 a day at a \$6.41 costs per click (due to 2nd price auction nature [156]). At Google’s suggested bid (\$0.24-\$0.56), we see roughly 15 to 20 clicks on average per day bringing the total daily costs anywhere from \$1.30 to \$1.60 for search engine marketing

Google allows advertisers such as Civil Good to run advertising campaigns by placing bids on search engine keywords in a generalized second-price auction. That is, “Google looks at the Ad Rank of the ad showing in the next position” and only charges “the lowest bid amount that would have beaten that ad’s Ad Rank, rounding up to the nearest cent” [155]. Because advertisers pay what the next-highest advertiser bids, cost estimates can vary greatly. For example, Figure 1 shows that if a website were to bid at the maximum costs per click (CPC) needed to be at the top of every search related to online discussion keywords (\$6.41), costs could range anywhere from \$0 to \$101. A better approach would be to stay within Google’s suggested bid range of \$0.24 to \$0.56, which has much more predictable costs as outlined in the Figures 1 and 2.

We can thus expect anywhere between \$1.30-\$3.40 in ad spend cost per day on search engine marketing rounding out to \$39-\$102 in monthly costs for SEM as shown in Figure 3.

Display advertising is advertising on other sites. In addition to AdWords, Google also runs an extensive Display Network and provides a similar tool for es-

Figure 1: High-End Cost Per Click (CPC) Details



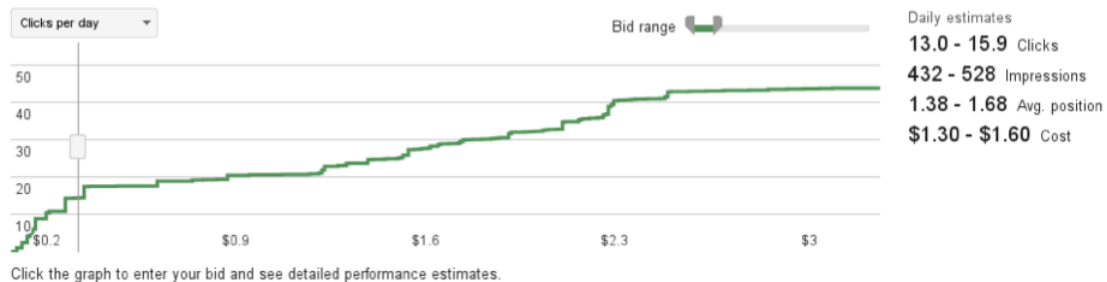
Clicks vs. cost graph from Google AdWords Keyword Planner for maximum CPC. Details the range of estimated costs for the high end of the curve.

Figure 2: High-End Recommended Bid Data



Clicks vs. cost graph from Google AdWords Keyword Planner for high-end recommended suggested bid (\$0.56). Estimated number of impressions and clicks, average position, and cost displayed on the right.

Figure 3: Low-End Recommended Bid Data



Clicks vs. cost graph from Google AdWords Keyword Planner for low-end recommended suggested bid (\$0.24). Estimated number of impressions and clicks, average position, and cost displayed on the right.

Table 6: Google Display Planner Data

website	Relevance	Historical CPC	Impr./Week
pipmedia.com	10/10	\$0.00-\$1.00	20M-25M
wmd.com	9/10	\$0.00-\$1.00	20M-25M
salon.com	9/10	\$0.00-\$1.00	30M-35M
talkingpointsmemo.com	8/10	\$0.00-\$1.00	20M-25M
townhall.com	8/10	\$0.00-\$1.00	20M-25M
alternet.org	8/10	\$0.00-\$1.00	5M-10M
Total	-	\$0.00-\$1.00	100M-150M

Historical CPC and estimated impression weekly volume for top matching websites related to online discussion keywords.

timating display costs. This tool suggests relevant sites and helps estimate basic demographics and impression volume based on the provided options. In general, display advertising would provide much more volume, but these users will likely not have high intent to stick around and monetize. This is because display advertisements are typically designed to match user interests with a product rather than capture intent to purchase as SEM does [93]. To determine the sites best suited for running the display campaigns of a website like Civil Good, Google’s Display Planner tool asks for general topics that might pique the interest of the site’s users. The topics for this case study were “online discussion” “online debate” “discourse” and “politics.”

As shown in Table 6, most terms related to online discussion or discourse have historically had low costs per click (CPCs). Most of the placement sites have relatively low number of impressions per week (compared with pre-built ad groups), but also have extremely low CPCs. Assuming a CTR equal to the national U.S. average (0.10 percent), we can get roughly 100,000 clicks per week at 100 million impressions at costs of less than 1 dollar per click [61].

Display advertising costs are difficult to predict as the historical CPC has ranged anywhere from \$0 to \$1. At a volume of 100,000 clicks per week, monthly

costs can be anywhere from \$0 to \$400,000.

Additionally, advertising on sites like Facebook seems to be more about just creating a social presence and actively engaging users rather than running display campaigns [126]. Most of the cost would likely come from hiring someone to manage and update these social media accounts as building a Facebook Page for people to ‘like’ is free of charge. The manner in which display advertisements are created depends on whether or not the advertiser elects to set up a Page. With a Page, there are several different options: getting more page likes, promoting page posts, and bidding for CPC. If the advertiser does not have a page, there are still plenty of options to choose the target audience. The audience can be chosen based on what people “Like,” right down to the precise interests such as hashtags for “Politics,” “Online discussion,” “People,” “Society,” and “Being Nice.”

CPC estimates are similar to that of regular display advertising networks, but advertisers have much more control over the target audience. If Civil Good were to run advertisements as a form of revenue, it would be very beneficial to carefully study how sites like Facebook utilize existing social information run display advertisements. Facebook recommends that bids be placed between \$0.46 and \$0.93 for a campaign run with the aforementioned interest keywords. The targeted audience would include 18,200,000 users.

Other social media marketing tactics such as Twitter and Instagram involve interacting with users at the individual level. They are usually used to connect more directly with users and to advertise promotions for existing sites. The usage of this method of advertisement should not be used until a website reaches a critical mass of users and needs to increase user retention.

Email marketing involves sending a message about a service or product directly to a customer using email. It is considered highly efficient because it has

“high response rates and low costs,” which are “rapidly turning email marketing into an invaluable tool” [309].

For existing users, email marketing can be used to send reminders of current activity on the website, remind users to return and complete their existing conversations, or solicit donations. This reminder could encourage a user to revisit the Civil Good website and aid with user retention. However, in order for emailing existing users, the users must provide their email addresses. Since Civil Good does not mandate that an email be specified during the sign-up process, there may very well be user accounts that do not specify any email, which can drive down the reach of an email marketing campaign. From the aspect of privacy, since email addresses are not linked to the users’ account, there is no personally identifiable information that can be used to connect a user with their Civil Good identity.

The other important aspect is whether email advertisements are only sent to users who have “opted-in” to receiving them. If the user has given their consent to receiving marketing information, there is a “higher response rate” for email marketing and the consent also helps to “develop the relationship between the consumer and the company” [350].

While the broader concept of email marketing also involves emailing a purchased list of potential customers’ emails, we do not recommend this approach since this is an example of email advertising that is not opt-in. Since the potential customer has not consented to receiving these emails, they cross the gray area into being considered “spam.” Since spam messages are unwanted, they can easily convey a negative impression of Civil Good [266].

### **12.3.2 Conclusion**

While both general and social display advertising provide a much wider reach at a lower cost per click, the estimated costs are still much higher due to the increased volume. The users brought in by display advertising may not necessarily be as intentful to stay and join the community and are probably less valuable users from a monetization perspective. For the purposes of retaining users, email marketing has the potential to be a useful option. It is recommended that Civil Good pursue search engine marketing and email marketing but not display advertising.

## **12.4 Revenue**

Websites that grow large often utilize a variety of different means of monetizing their content to support their operating costs. Viable options to monetize Civil Good include targeted display advertising, sponsored affiliate posts, email remarketing, premium user subscriptions, donations and selling user data trends. Each method has its own set of advantages and disadvantages. All of these options become viable once the site gathers enough regular users and visitors [394]. The overarching question when considering each method then becomes: does this method significantly diminish user retention rates? If so, does the method's profitability warrant its continued use?

### **12.4.1 Background and Discussion**

Although display advertising has its pitfalls regarding user retention and privacy, it remains "one of the most common forms of website monetization" due to both its uncomplicated implementation and its high efficacy [204]. Much of the potential value in display advertising stems from the idea of demographic-based targeted

display advertising. Civil Good can uniquely provide very fine-grained and detailed demographic information about its individual users, which advertisers would find very valuable. “PACs,” for example, “would definitely pay a lot of money for that particular demographic targeting” [394]. Unfortunately, advertising is not without its downsides. Some studies show that “the mere existence of ads decreases retention of both site and ad content” as users can lose interest at the just the sight of an advertisement [261]. However, other studies show that users’ views and perspectives on advertising can vary significantly from person to person. They show that “different types of people view advertising differently, and on different dimensions than conventional demographic or usership criteria” [265]. Additionally, the same person will have a different response to advertising depending on their motive for Internet use (e.g. research, communication, or entertainment) [352]. Any attempts to judge how advertising would affect the typical user would be ineffectual until the actual launch of the site. Techniques exist to mitigate the effect advertising has on user retention (like using inline ads rather than pop-up ones), but little can be done to eliminate the effect in its entirety [261]. Users in a research study in the *Journal of Advertising* rated pop-up advertisements extremely low, and skyscrapers (vertical banners) were the most favored format [50].

However, the Civil Good founder notes that user privacy could still be a concern [252]. On the other hand, ignoring display advertising altogether would be cutting off what marketing expert Eugene Suzuki considers “really valuable about this business... the very very focused ad targeting” [394]. Although display advertising can often be shown to drive some users away, its lucrativeness almost definitively justifies the potential loss of users.

Advertisements can be included on Civil Good in a manner that neither obtrudes on the user experience nor violates any user privacy. Advertisements can

be deliberately chosen such that they blend well with the site design, which can lead to increased click-through rates and increase potential earning [238]. Civil Good could also utilize text-only advertisements, which are very subtle and appear to be just another link on the page [178]. These types of advertisements tend to be much less obtrusive and often come with lower costs. A 2008 study by Mike Vorhouse of *Ad Age* reports that although users might pay to remove advertisements, most users are not likely to pay at all, and no user is willing to pay more than \$30 a year [419].

Certain advertising platforms exist such as Project Wonderful that deal with common advertising problems such as click fraud and lack of choice on what advertisements are displayed [439]. The online community reddit is a good case study of a site that utilizes unobtrusive ads and target users with specific interests without compromising user privacy [372]. In addition to traditional display advertisements, reddit has self-serve advertisements that integrate with site content [6]. Since one of reddit's primary features is sharing links to interesting content on the Internet, reddit allows advertisers to pay for links to their content to be promoted to a wider audience by leveraging the existing structure of the site [6]. reddit also allows advertisers to target communities called "subreddits," which focus on a particular topic or set of interests [372]. Their advertising platform, Adzerk, allows users to vote on whether they like the advertisement, which affects visibility for users with similar interests [355].

Affiliate marketing is another method similar to display advertising wherein a website develops a partnership with the affiliate company. If the user clicks on the ad and monetizes for the affiliate company, a portion of the revenue goes back to the website. This method monetizes best on a product-centric site, but it can also be effective on sites like Civil Good that facilitate discussion [204]. However, there does exist a tradeoff between the efficacy of affiliate traffic and the level of



trust users have on the site’s objectivity [204]. Civil Good could really benefit from targeted affiliate advertisements. However, for a site like Civil Good, user trust in the site’s objectivity is key to user retention, and targeted advertisements could potentially erode that trust. Aside from user trust, it would be difficult to form an affiliate partnership while continuing to protect user privacy at the expected level. Users would likely be unable to use the site without concerns about privacy.

Some websites such as reddit and Pandora generate some revenue through optional, paid online subscriptions [342, 282]. Users could pay to remove advertisements, beta test new features, get access to an exclusive members-only forum, and more [342]. This subscription model, according to Michael Johnston at Monetize Pros, “can be a great opportunity to open up a lucrative revenue stream” provided that users consider the content on a specific site valuable [204]. However, this business model is not very profitable. In 2011, Pandora experienced a net loss of 6.8 million dollars even after doubling its revenue from the year before [282]. reddit by 2010 had only 6,000 of its eight million users subscribe to reddit gold, and the fact of the matter is that reddit is still not profitable [213, 343]. It should also be noted that if display advertisement turns out to be even more successful, having subscription-based ad removal would be “putting content behind a paywall” meaning ads get “fewer page views and lower ad revenue” [204]. The bottom line is that even after having obtained aforementioned “critical mass” of users who all believe that a given website provides a service worth supporting, volunteer-based online subscriptions are not enough to break even with operational costs [394].

For a site like Civil Good, donations and company grants are another avenue with strong potential. Online donations have been increasing in volume in the last decade from just 4 percent of people giving online in 2002 to 65 percent in 2012 [140]. Blackbaud reports that online donations in the United States grew 10.7% in 2012

on a year-over-year basis, and that 7% of all nonprofit fundraising came from online giving [251]. However, it is important to note that these statistics mostly pertain to charitable nonprofits rather than online services like Civil Good. Donations work without needing to achieve a critical mass of users.

The main disadvantage with donations is that donors have no obligation and little incentive to provide continual support. Donation amounts can be affected by a number of factors, including what information is made available to donors, clarity of the site's mission, reputation, and organizational size [410]. The sheer quantity of contributing factors makes it difficult to create an accurate prediction model for this method's efficacy. A very rough estimate of the potential revenue in donations can be built from the statistics of other nonprofit sites such as Wikipedia. Wikipedia's average donation rate was 0.302 percent of users with an average donation of \$19.05. At those rates, a modest estimate of 1,000 monthly users would generate roughly \$57 in monthly revenue [314].

The demographic who is most likely to donate online include younger donors with little to no prior history donating through traditional means (such as a check or credit card) who have lived in their residences longer. A study from Blackbaud concluded that the most significant driving factors in converting online donors include age, giving history, and length of residence [356]. Other interesting factors from this study conclude with marginal significance that more online donors come from households that include two or more adults, individuals with an income level of \$50,000 or more, verified homeowners with a home value of at least \$175,000, individuals who identify as male and credit card owners [356]. Of these factors, Civil Good could reasonably use information provided by users on estimated income, current home value, gender, and homeownership status to affect the interval during which donations are solicited. For example, individuals with a reported income of

less than \$50,000 could be solicited less frequently than those with higher reported income.

Another important question when soliciting donations is the best time of year. 30% of online giving happens within the final three months of the year, with a majority of online fundraising taking place during the final weeks of December [250]. Blackbaud's research indicates that after six solicitations a year, the likelihood for long-term loyalty diminishes significantly [356]. Therefore, it is recommended that Civil Good record in its database the number of times that a user is specifically asked to donate and that the number not exceed six times a year.

Community loyalty to the site can also aid in increasing donations. For example, in October 2013, the online community reddit experienced a sudden, large increase in financial contributions via their premium subscription service (reddit gold) after one user created a post that informed the rest of the community that reddit has been unprofitable despite its large user base [109, 112]. Creating this sense of community loyalty will become very important in securing long-term financial sustainability. In response to this overwhelming support and because the community requested it, reddit started to display a "daily reddit gold goal thermometer" [112]. This thermometer functions as a progress bar for reddit's daily fundraising goals. Once Civil Good has established a large, loyal community of users, it may be beneficial to include a similar thermometer so that users can be more aware of the site's financial stability.

Another proven solution is building an online store to sell physical merchandise directly to users. Research shows that online stores work best for users who are loyal and want to support the main site, but they are bad for transient audiences. "Power users," as they are called, would most likely be want to purchase some branded merchandise [204]. DFTBA Records, a record label focused on supporting

the music and merchandise for content creators on the community site YouTube, had sold over a million dollars in merchandise [165]. For Civil Good, we need to analyze whether people would want to stay on long term, which could only be done after Civil Good is implemented. As with other methods, it should be profitable provided we have a “critical mass” of users [394]. Users can maintain their privacy because it is separate from their user account and would probably utilize third party services such as Braintree, Authorize.net, or Google Checkout [99]. Having an online store does not hurt user retention and would likely elicit a positive response.

Civil Good could also monetize by selling user data for data mining opportunities. Many other companies including MasterCard, American Express, eBay, Twitter and Foursquare are already in the business of selling user data, especially to advertisers [113, 108, 262]. MasterCard, for example, makes its data available through its Exelate service, which feeds anonymized transaction information to digital ad exchanges for targeting [211]. The information is organized by ZIP code and cannot be traced to individual users [113, 211]. Location-based social network Foursquare works with advertising company Turn to utilize its extensive database of user location data to retarget Foursquare users on other websites [262]. Location information is considered to be the most prized data for ad targeting [262]. Twitter disclosed that they earned \$47.5 million in revenue from selling off data to companies that analyze it for insight into news events and general trends [108]. Other companies, such as eBay, provide information as a service in a less direct manner. Instead of releasing raw data, eBay has internal analytic teams working to analyze search, buying and selling trends to produce a service called The Inside Source [296]. The Inside Source serves as a “cultural barometer” that purports to predict trends in fashion and shopping using eBay’s proprietary data, and is provided free of charge for end users [296, 30]. The online auction company profits when those who follow

The Inside Source click on links to purchase trendy items on eBay. With some services such as Handshake, users volunteer and negotiate the price of their personal information and sell it directly to interested companies [58, 216].

With a large enough dataset, selling user data to advertisers can be immensely profitable. As shown by MasterCard and eBay, the information does not have to be directly linked to individual users and can be anonymized and grouped. Each company takes advantage of their most valuable information. MasterCard sell its transaction information, Foursquare has its user location data, eBay has its search and transaction information, and Civil Good has its demographics and poll information. For example, Civil Good could sell information to electoral campaigns and political action committees (PACs) about correlations between political party affiliations and popular debate topics. These PACs could use the information to advise their candidates to uphold a popular opinion in public debates and win a majority of votes. Alternatively, Civil Good could selectively make available raw data for other companies to perform their own data mining and trend analysis.

#### **12.4.2 Conclusion**

Targeted display advertising is easy and very lucrative. However, if done incorrectly, advertising can create privacy concerns for users. To mitigate these concerns, it is recommended that Civil Good utilize advertising platforms such as Adzerk and Project Wonderful, which would give Civil Good control over what advertisements are displayed. Civil Good should use either text-based advertisements or skyscraper display banners that blend with the site design to make the advertisements less obtrusive, and Civil Good can elect not to include advertisements in the conversation Web pages. In order to determine whether to implement certain methods such as an online store, the practical manner would be to analyze the user base once the website

is already functional. Donations should be solicited no more than six times a year and should utilize demographic information such as age and gender to determine the frequency with which to ask users for donations. Ultimately, a website like Civil Good needs to build a large enough and loyal user base before it can expect to see any reliable source of income.

## 13 User Interaction

We examined how to simplify and streamline various user interaction scenarios. We focused on improving usability and reducing the learning curve.

We also researched and developed a rating system for conversations. We contrasted this rating system with the original rating system provided in the Civil Good design document [252].

We designed static mockups of the website, with one version based on the Civil Good design document, and the other version incorporating our recommended changes to the user interaction scenarios. We then conducted a usability study using these two mockups to determine the effect that our recommended changes had on the website's usability.

### 13.1 Related Work

The user interaction research by and large examines ease of use versus the functional benefits potentially to be gained from a complex user interface, and discusses user reactions arising from those issues. It also examines some aspects of how users interact with the site.

#### 13.1.1 Simplicity versus Complexity

The general concept of simplicity versus complexity in website design was examined by Carlos, Raquel and Carlos [56], who performed an analysis of the website design of Amazon.com and Easyjet.com. The authors first noted studies that emphasized simplicity and ease of use as important factors in website success:

- Jarvenpaa and Todd found that the effort required versus the convenience to be an important factor affecting users' attitudes towards the website [200].

- Lohse and Spiller found ease of navigation to be an important determinant in increasing website use [245].
- Liu and Arnett found “ease of use of the system” to be one of the four major determinants of a successful Web design [243].

Focusing on the factors of simplicity and ease of use, the analysis of Amazon.com found that its simplicity of navigation, flexibility of the search engine (for example, permitting both a general search and a more detailed search), and efficiency of the purchasing process (which can be completed in “four simple stages,” or by an optional “one-click to purchase”) to be major determinants of the site’s success. The analysis of Easyjet.com found the ease of the shopping process, which is comprised of a search function and a small number of navigation steps, to be an important factor in the usability of the site. Based on these observations, the authors recommended that website design incorporate “navigation characterized by simplicity” [56].

Bernard [37] studied various human factor concerns in the design of websites, such as information positioning, arrangement of menus, navigation structure, text presentation, image use, frame use, interface usability, user annoyances, age accessibility, and conformance with Web conventions. Many of these impact the issue of simplicity versus complexity. Bernard recommended limiting the amount of information presented at any one time to avoid the “art museum” problem of overwhelming the user.

Bernard emphasizes the desirability of simplicity in Web design in multiple areas, including structural and informational areas, stating that “the goal should always be to reduce the complexity of the site as much as possible.” This principle applies to both interface elements (Bernard quotes Mies van der Rohe, “less is more”)



and navigational processes.

Consistent with these findings of the importance of simplicity in Web design, Iyengar and Lepper [197] found that while choice is preferable to no choice, human ability to manage choice is limited, and that there is a limit to the number of options that can be efficiently processed. More specifically, Iyengar and Lepper found that people perform better and receive more satisfaction in tasks involving a limited array of 6 choices, as opposed to an array of 24 or 30 choices. The authors also noted the following studies with similar results:

- Dhar discussed conflict caused by a multiplicity of choices, often resulting in a decision to not make a choice [100].
- Shafir, Simonson, and Tversky examined the increase in individuals deferring decisions as additional choices are added to the option set [368].
- Timmermans found that individuals processed a smaller percentage of available information in making choices when the number of available alternatives increased from three to six, and then to nine [406].

Based on those studies and their own results, Iyengar and Lepper concluded that the provision of too many choices causes “choice overload,” which demotivates the individual from completing the task, or from completing it properly, causes individuals to experience conflict and difficulty valuing the many options, and inhibits users from engaging in a process appropriately, and completing it.

### **13.1.2 Forced Response and Restrictive Navigation**

The requirement that a user input a significant amount of information to properly participate in the site’s processes was studied by Stieger, Reips, and Voracek [387],

who found that “forced-response” questionnaires (online forms that require users to provide responses before being allowed to progress) increase dropout rates and negative behaviors. The authors observed behaviors such as skipping questions, providing random responses in order to complete the form faster, and ceasing to use the website altogether.

Dailey [87] defines website navigational cues as items aiding site navigation, such as image, text, and button links. A restrictive navigational cue is one that limits a user’s freedom of navigation, such as a page with only “previous” and “next” links. Dailey cited Brehm’s [44] theory that negative behaviors are caused by restrictions of an individual’s behavior when those restrictions are contrary to the individual’s expectations. Dailey then generalized that theory to Web navigation, hypothesizing that users’ experiences on the World Wide Web with websites that are navigationally unrestrictive create expectations of unrestricted navigation in other websites. Restrictive navigational cues on other websites will then be contrary to those expectations, causing negative reactions such as an increase in negative attitudes and emotions towards the website, in addition to an increase in avoidance behaviors [87].

In our interview with Professor Djasmasbi, she stated that “I think optional is a good idea...you don’t want to force [users] to do anything” [102].

### **13.1.3 Anonymity**

Tsikerdekis [412] studied the effects of anonymity on online behavior. While the most commonly perceived downside of anonymity is presumably its promotion of aggressive behavior, Tsikerdekis conducted surveys on controversial topics with social media users, and found that there was equal or less aggression by anonymous participants than by those using a pseudonym. Tsikerdekis also cited studies show-

ing anonymity to have positive effects:

- Ainsworth, Gelmini-Hornsby, Threapleton, Crook, O'Malley, and M. Buda found that anonymous voting reduces the impact of “groupthink” (a psychological process within a group where decision making is adversely influence by the desire for conformity) and promotes individuality in decision making [15].
- Lea, Spears, and de Groot found that anonymity reduces concerns about others’ judgments, promoting a more objective focus on the task at hand [229].

Tsikerdekis thus stated that “complete anonymity is recommended for communities where debates take place regularly, and users that feel they have strong opinions about topics can be found in a big percentage of the population.”

While some people argue that anonymity is a strong promoter of “trolling” or “flaming” (abusive or malicious comments), the opposite was shown in South Korea in 2007. South Korea’s government repealed a previously established policy that all websites with over 100,000 viewers require the use of real names. The policy was found to reduce malicious comments by only 0.09% [64].

Lapidot-Lefler and Barak [223] found that the most important factor in promoting online disinhibition is the lack of eye contact, as opposed to the anonymity the Internet can provide. Lapidot-Lefler and Barak noted a study by Douglas and McGarty [104] showing that users whose identities were revealed engaged in more aggressive online behavior than anonymous participants did, demonstrating that anonymity alone is not the strongest determinant of aggressive social media behavior.

#### **13.1.4 Gamification**

Gamification is defined as “the use of game design elements in non-game contexts” to motivate desired behaviors and increase user engagement and motivation in the user’s interaction with a structure, such as a social media website. Typical gamification methods involve establishing rewards (such as points, badges, or achievements) and ranking systems (leaderboards) to increase users’ engagement in the structure. The mechanisms underlying the efficacy of gamification are social psychological processes providing desired intangible rewards, such as social approval by others and group identification [96]. A classic example of effective gamification intangible rewards is merit badges earned by Boy Scouts.

It is generally accepted that gamification can serve to increase user involvement in the use of social media websites, as demonstrated by Cetin [60] with regards to the SAP Community Network. The SAP Community Network is a social network for SAP employees to discuss products and technology. Using missions, badges, points, and a leveling system, the developers increased general activity by 400%, and point-generating activity by 2,210% in just one month. There is empirical evidence supporting gamification’s positive effects on engagement and motivation, as Jennifer Thom, David R. Millen, Joan DiMicco [403] found that adding gamification elements to an enterprise social networking system dramatically improved content contribution. When the gamification elements were later removed from the system, user activity significantly decreased.

### **13.2 Recommended Changes to Website Design**

Given the potential complexity of user interaction with Civil Good, the overarching principle applied to the below topics is to attempt to simplify the various stages of

interfacing with the website without losing the depth of interaction that complexity can potentially provide. Civil Good will have to monitor user interactions and continually tweak the interface as it grows to strike the proper balance between ease of use and functional depth.

### **13.2.1 Facts, Topics, and Opinions**

The proposed Civil Good signup process requires the user to enter a significant amount of information [252], such as:

- Age, gender, and education
- General occupation, skills, and industry
- Marital and family status
- Military status
- Languages known
- Religious affiliation
- Political party affiliation
- Approximate income and wealth
- Myers-Briggs classification
- Opinions on current high-profile issues

We believe that this proposed signup process is overly complex and restrictive, and therefore likely to result in negative reactions in many potential Civil Good users, such as abandoning the signup process or entering random and incorrect information. Such negative behaviors were observed by Stieger, Relps, and Voracek in

connection with similar “forced-response” questionnaires [387]. Users accustomed to quick and easy signup processes of other websites, but who are otherwise ideal users for Civil Good, will be put off by the long signup process proposed for Civil Good, as observed by Dailey. Additionally, the complexity of the proposed initial signup process is likely to cause “choice overload” as observed by Iyengar and Lepper, which would demotivate potential users from completing the signup process properly.

Our recommendation is that users be presented with a simple and streamlined signup process that only requires the bare minimum of information. They should not be asked to provide facts or opinions during the signup process; instead they should be required to provide information about themselves during the process of creating a Seek. When a user creates a Seek that involves criteria that they have not yet provided about themselves, the user should then be prompted to provide this information. This allows the site to ask the user for small amounts of information at a time, as opposed to an initial large amount. Our recommendation is consistent with the findings of Bernard and Carlos, Raquel, and Carlos, both of whom emphasized the importance of simplicity in website design [37, 56], supporting the proposition that a simplified signup process will decrease negative attitudes and will increase website usability and efficiency.

It is true that a simplified signup process increases the ease of a user creating a “throw-away” account that could be used for improper purposes, such as trolling or spamming. However, because a user will be required to enter facts and opinions to either create a Seek or be matched to a Seek, genuine users can be differentiated from “throw-away” users by the quantity of facts and opinions the user has provided, measured against the length of time the account has existed. An account that exists for a pre-determined amount of time without the provision of facts and opinions can

be flagged and, if necessary, terminated.

Moreover, the addition of gamification factors to accounts adds value, such that a user will incur a loss from the cancellation of their account. Genuine users of Civil Good will want to retain their accounts to keep whatever benefits they have accrued from the gamification factors, rather than serially creating new accounts.

Once a user has obtained an account, they will be prompted to enter facts and opinions in order to create a Seek and to be eligible to be matched to a Seek. The site could reward the provision of facts and opinions with gamification benefits. The site should be designed to encourage active participation by providing no benefit to lurking behaviors. If a user does not engage in conversations (which require first providing facts and opinions), there will be few options for activities within the site.

While requiring a user to provide more than minimal information at sign-up may be considered as Civil Good grows, the foregoing considerations tilt towards a more simplified sign-up process at the beginning, to help grow the user base. A rapid growth of the user base while the site is still young can only be beneficial to the success of Civil Good, and an overly complex sign-up process would ward off many potential users, slowing the growth of the number of users.

### **13.2.2 Polls and Surveys**

The current Civil Good design includes mandatory periodic polls and surveys in order for a user “to remain active” [252]. We recommend that polls be optional and include a “Skip for now” button.

Requiring users to complete a poll in order to remain active is a prime example of a forced response questionnaire, which Stieger, Relps, and Voracek have shown is linked to increased dropout rates and negative behaviors, such as users entering false information [387]. The requirement for users to complete a poll could

also be seen as a restrictive navigational cue that reduces a user’s control over use and navigation of the website. These have also been found to cause bad results in users, such as negative emotions and negative attitudes about the website [87].

It is true that making polls optional may promote a volunteer or self-selection bias, whereby users with certain opinions on an issue may be more likely to respond to the poll, thus skewing the poll towards the attitudes of that self-selected group. However, because poll results are not intended to be scientific and results are not represented to reflect the views of any designated population [252], volunteer bias in polls is viewed as less of a negative than the significant adverse effects that would be caused by mandatory polls.

Therefore, we recommend that polls should not be mandatory and should not be required as part of the sign-in process. Polls must be optional. They can still be presented as part of the sign-in process and periodically during use of Civil Good, but an escape route must be available. Instructions on how to return to the poll at a later time should be included. However, the site should offer gamification benefits to those who complete polls, in order to encourage their use. These benefits could be provided on an ascending basis, so that greater benefits are provided for completing more Polls.

### **13.2.3 Member Identity**

The current design proposes the use of random “dynamic” usernames to identify users during conversations and other Civil Good activities [252]. Our recommendation is that names should not be shown in a conversation, making conversations completely anonymous.

Requiring users to participate in conversations anonymously will improve the quality of conversations between users. Anonymity will reduce disinhibition and en-



courage users to voice their true opinions, as shown by Tsikerdekis [412]. Based upon the findings of Lapidot-Lefler and Barak that the most important factor promoting online disinhibition is lack of eye contact (as opposed to mere anonymity) [223], anonymity is not viewed as a strong promoter of malicious behavior in Civil Good. Civil Good as presently designed does not include support for Webcams in conversations. This finding is consistent with the failed South Korean experience where requiring websites with over 100,000 viewers to use real names reduced malicious comments by only 0.09% [64].

Finally, because Civil Good will be employing leaderboards showing members who have achieved the best ratings in various categories in order to take advantage of the positive effects of “gamification,” anonymous participation in conversations will serve to moderate the negative effects that accompany specific users becoming “celebrities” within a social community. The existence of such “celebrities” may cause discussions and ratings to be more attuned to the identity of the participant than their conversation performance. Users may react abnormally to the opinions of a “celebrity,” becoming more attuned to the celebrity’s opinions rather than engaging in an honest debate. Having complete anonymity will be useful in avoiding these issues.

Logistically, the conversation window should refer to a user’s conversation partner as “Partner,” with no further identification. The UI should permit a user to add a descriptive tag to a conversation to enable the user to keep track of multiple conversations he/she may have ongoing.

#### **13.2.4 Faving**

The Civil Good design proposes that following a conversation, users can request that the other user become one of three types of faves: a Fave for Conversing, a

Fave for Viewing or a Full Reveal Fave [252]. We recommend there be only one type of Fave. If two Faves later engage in a conversation with each other, they cannot give each other a rating.

The original Fave proposal is overly complex and therefore demotivational, as it causes users to experience conflict and difficulty valuing the options (“choice overload”), as found by Iyengar and Lepper. At the end of a conversation, users should be faced with the simple choice of agreeing to be a Fave or not, with the consequences of Faving clearly defined. Therefore, we propose only one type of Fave, which would permit users to:

- See each other’s display names
- Start conversations with each other directly
- View each other’s conversation histories

The recommendation of only one type of Fave is consistent with the emphasis on simplicity of Bernard and Carlos, Raquel and Carlos. When a conversation has ended, a yes/no Fave option, which clearly defines the meaning of agreeing to Fave status, will be more easily understood and utilized by site users.

As originally proposed with Full Reveal Faves, we agree that Faves should not be permitted to rate each other. The number of Faves a user accumulates will itself signal the quality of a user’s interactions, and we want to avoid users giving their Faves overly high ratings.

It is hoped that a Civil Good user who has completed a conversation will want to engage in further conversations, and the post-conversation process should not hinder that goal by being burdensome. We recommend a short and simple post-conversation Fave option process, to permit the user to easily return to the heart of Civil Good, the next conversation.

### **13.2.5 Sending a Seek**

Once a Seek is created, Civil Good disseminates the Seek to other members to find members matching the Seek parameters who are willing to engage in a discussion of the subject of the Seek. The design proposes that the Seek first be sent to a subset of members and, if necessary, later be sent to additional members [252]. We recommend that seek requests should be sent to all members who match the Seek criteria, rather than a limited set of members at a time.

At Civil Good’s inception, and as Civil Good works to grow membership and encourage participation, it would be counterproductive to limit the transmission of Seek requests to a limited set of members. Certainly at the outset, Civil Good should permit all members to view all matching Seeks in order to maximize user to user interaction and the number of conversations that can occur. Limited publication of Seek requests may be useful as the community grows, in order to avoid the motivational effects of “choice overload,” as discussed by Iyengar and Lepper [197].

Civil Good should continuously evaluate the total number of Seek requests, the number an average user sees, and the number of conversations taking place, in order to evaluate if and when the transmission of Seek requests should be limited.

### **13.2.6 Specifying Seek Criteria**

The originally proposed design requires users to input various Seek criteria to be used in finding potential matches for conversations. If an insufficient number of matches are found, the Civil Good engine might reject the Seek, force the conversation to be private, or require the member to modify the criteria [252]. We recommend that when a Seek is determined to be too narrow, such that too few matches are found, the Seek instead be allowed and the Seek criteria automatically relaxed until there are enough matches. Upon an acceptance, the user who created the Seek should be

informed of the search terms that were not matched.

The original proposal for handling overly narrow/specific Seeks is complex and has the potential to frustrate users. Civil Good should not “reject” a Seek, as this will discourage users, who may not understand why the Seek was rejected. Bernard referenced difficulty in finding what one is looking for as one of the top five website annoyances [37], a category in which the rejection of Seeks would fit.

Similarly, Civil Good should not require the user to modify the Seek criteria, as that restrictive navigational cue may cause a cycle of modification and rejection, which would frustrate and discourage users. A system where the Seek criteria are relaxed automatically will also encourage users to learn to fashion appropriate and sufficiently broad Seeks, since the user is informed about which criteria were too constraining.

The option of forcing the conversation to be private is too restrictive to be practical, as Civil Good should always give users the option of whether or not a conversation should be private. Relaxing the Seek criteria permits this flexibility and does not overly restrict users, which would cause frustration and annoyance.

In order to prevent a user from locating another particular user by creating a very narrowly defined Seek, Seeks should automatically be expanded until a set minimum number of matches are found. Users should be told which aspects of the Seek were expanded, but not in sufficient detail to permit the user to devise a work around of the anonymity requirements and locate a particular user. The requirement of a minimum number of matches to a Seek will also prevent a user from iteratively modifying a Seek to “investigate” a particular user’s beliefs and opinions.

This design should also deter users from using Civil Good for illicit activities and taking advantage of the privacy features offered by the site, as the effort required to find a partner with whom to engage in illicit activities will be substantial,

compared to alternative methods currently available in society and on the Internet.

### 13.2.7 Creating a Seek

The proposed design requires that when creating a Seek a user decide whether the conversation should be interactive (similar to instant messaging) or non-interactive (similar to email). The design also requires users to decide whether a Seek should be advertised as an “Instant Match” [252]. We recommend that all conversations be interactive, and that the Instant Match option be eliminated.

Civil Good’s Seek creation process is quite complex. The Civil Good proposal for creating a Seek has (among others) the following options for a user to consider:

- Topic
- Type of conversation (Debate, Query, Devil’s Advocate, Whines/Gabs, or Rant)
- The user’s existing beliefs about the Topic
- Whether the conversation will be interactive or non-interactive
- Whether the Seek will be an Instant Match or not
- The parameters of the ideal partner (religion, gender, opinions, history, ratings)
- Whether the conversation will be Private, Members-Only, or Public for viewing after it finishes

This is very intricate and is likely to cause demotivating “choice overload” in many users, as found by Iyengar and Lepper. Simplifying the Seek creation process is also in line with the emphasis on simplicity and efficiency of use of Carlos, Raquel

and Carlos. Therefore we recommend that the Seek creation process be winnowed down as much as possible to its essentials, and stripped of all but the most vital options. Two aspects that should be eliminated are the interactive/non-interactive choice and the Instant Match option.

All conversations in Civil Good should be of the interactive/instant messaging type if both users are logged on and available. This is how popular modern social media chat systems work, including Facebook chat, Skype, and text messaging. Similarly to those systems, if one of the users is not currently logged on, then messages sent by the other are queued up so that the user can see them once they log back in. This allows for quick back-and-forth conversation in some cases, but also allows for users to leave and return to the conversation later. It additionally allows users to take their time in crafting a message should they so desire. Introducing the additional choice of completely non-interactive conversations needlessly complicates the Seek process. It is also unnecessary, as interactive conversations can be non-interactive if both users decide to take their time in sending messages to each other.

There also is no need for the Instant Match option. For Civil Good to function in a satisfactory manner, the normal Seek process must be sufficiently fast so that users are not frustrated by the delay in learning whether there is a match for their Seek. Given that necessity, the Instant Match option is unnecessary and so we recommend that it be removed.

### **13.2.8 Gamification**

Civil Good should continuously strive to increase both the quality and quantity of user interactions on the site. Gamification has been shown to increase both the quantity and the quality of content on social media websites [60, 403]. Therefore we

recommend that gamification elements should be utilized to increase user engagement on Civil Good.

Conversations will be subject to user ratings in specific areas relating to the quality of the conversation [252]. Users could also be rated in areas relating to their level of activity, such as most active, most sought after, or most conversations rated. Leaderboards reflecting conversation ratings were already planned to be posted on the site [252], and these could be expanded to cover additional ratings. Professor Djamasbi recommended that users be required to take part in more than one conversation before being placed on the leaderboards, as a way of encouraging initial user participation [102].

Gamification can interact with the polling process in two ways. Users can be rated on the number of polls they complete, thus encouraging poll completion. Polls could also be constructed to solicit user input on the optimal use of gamification structures. Civil Good should launch with a “starter set” of gamification elements, so that the site does not come off as too “gamey,” and polls should be used to evaluate user reaction to existing gamification elements and the potential addition of new gamification elements [102]. This idea is discussed further in Section 8.

Finally, Civil Good should ensure that gamification elements are used properly to reward users who are not only active but are making positive contributions to the goals of the site. Quantity of activity alone should not be overly rewarded, so that users cannot obtain leaderboard status solely via activity that does not further the quality of the site.

### **13.2.9 Visually Impaired Users**

In order to make Civil Good accessible to the largest possible population, the interface designers and developers should work to make the website accessible to persons

with disabilities [102]. Guidelines for accessible Web design have been published by the World Wide Web Consortium [421]. These guidelines should be reviewed by the designers and developers, who should strive to comply with them to the greatest extent possible.

#### **13.2.10 Conclusion**

The issues of complexity versus simplicity and restrictiveness versus freedom associated with a user's interactions with a social media website can dramatically affect the user's attitudes towards the site, activities when using the site, and the decision regarding whether or not to continue using the site. Civil Good should strive to provide users with less complex and less restrictive interface experiences, in order to avoid eliciting negative attitudes and behaviors towards the site. Civil Good should also make use of techniques designed to increase user engagement, such as gamification. By designing a website that is as attractive and engaging as possible, Civil Good can create an environment that will inspire users to take part in a larger number of higher quality conversations.

### **13.3 Conversation Page Improvements**

The Civil Good experience revolves around conversations. It is therefore important that the conversations page is well-designed and has features that will be helpful to users. The current features of the conversation page, as described in the original document, are outlined below, along with recommendations for new content to add and modifications to be made to existing content.



### 13.3.1 Background

The current model of the conversation page has users' chat messages appear inside a scrollable box. These messages are color-coded and positioned in such a way that it can be easily determined which message belongs to which participant in the conversation [252]. At the top of the conversation page, the type of conversation being held is displayed, along with the conversation partner's display name and a link to their profile page. The conversation page includes filtering options to sort prior messages in the conversation either chronologically (the default option) or by only the messages posted by a specific conversation participant [252]. At the bottom of the conversation page is a space for composing new messages, along with several formatting options. These options include bold-faced font, italics, underlining, the ability to change the size of the text, and also buttons to add a bullet point or an item in a numbered list [252]. The space where a message is composed includes a built-in spell-checker to help users compose more legible messages [252]. There is also an "object" button which allows a user to insert an image into the message, whereas audio and video files are not supported but can still be linked to [252]. There is an "undo" button, which will allow a user to reverse any change that has been made to the message. Once a message has been composed, the user is given the option to either post the message immediately or to post it later, meaning to save the message as a draft [252]. Lastly, at the bottom of the page is a button marked with a question mark that likely serves as a help menu or tutorial for how to utilize the various features of the conversation page [252].

It is possible for a user to alter a message that has already been posted. So long as no ratings have been given for the conversation, a user is allowed to edit any message that he or she has submitted during that conversation. In such an event, the user's conversation partner is able to see any edits that have been made

to previous posts [252]. If ratings have already been given for the conversation, the user is no longer allowed to edit previous posts, in order to preserve the integrity of the rating system [252]. The user is still allowed to post a follow-up message with any corrections or clarifications he or she wishes to make once ratings have been given. The exception to this rule is that a user may delete part of a sent message at any time, but cannot add to it [252]. This rule is meant to allow users to prevent personal information that they may have accidentally revealed from being publicly viewable even if the conversation is public. The user's conversation partner will still be able to search the revision history to find that deleted information, but no other user will be able to view it [252]. The conversation partner being given access to the revision history is meant to discourage users from abusing the ability to retroactively delete content from messages in an attempt to influence their ratings [252].

Two particular buttons of interest on the conversation page are "refer" and "source." The "source" button allows a user to highlight a portion of his or her message and have it point to a citation that appears on the side of the conversation page, outside of the scrollable box [252]. The "refer" button allows a user to highlight a portion of his or her own message and have part of a message sent by either participant appear as a reference on the side of the conversation page, outside of the scrollable box [252]. Any text that is highlighted by the "refer" or "source" buttons has an arrow pointing to the corresponding note outside the scrollable box, with a button included at the bottom of the scrollable box to hide all of these arrows if the user so chooses [252]. The "refer" button can be used in conjunction with the page's built-in search bar labelled "Find in Conversation" that allows a user to locate a specific claim or piece of information presented earlier in the conversation [252]. These two buttons, as well as the search function, make it easier for conversation participants to keep track of what each of them has said.

Additionally, it is possible for users to use the “source” option to not only provide citations to back up their claims, but to provide each other with sources for the purpose of additional research [252].

### **13.3.2 Discussion**

One thing to point out is that the “undo” button appears to be completely unnecessary from a logistical standpoint. If a user makes some sort of formatting mistake, he or she can still use the formatting buttons to correct whatever mistake that may have been. If the user accidentally deletes part of the message they were in the middle of composing, they can simply use their computer’s keyboard shortcut (the key sequence Ctrl-Z or Cmd-Z, depending on the computer) to undo the change instead of using the “undo” button. The “undo” button also does not serve to recall a message that might have been sent prematurely, nor would a user need it to. In such an event, the user can simply edit the message or post a follow-up message explaining the mistake depending on whether or not ratings have been given for the conversation by that point. There does not appear to be any situation in which the “undo” button would be useful.

Among the formatting options are two buttons: one that appears to increase the text size, and one that appears to decrease it [252]. In order to save space, it would be better to remove the two buttons and simply include a drop-down menu containing a list of text sizes for a user to choose from. The original design of the page does not tell the user what the current text size is, so a user has to judge it by eye to know if they have the text size where they want it after changing it. The downside to making a drop-down menu is that it increases the number of clicks a user has to make if they are only adjusting the text by one size. However, the original design does not specify the range of possible text sizes. If there are more

than three sizes available, and a user wants to adjust the text by more than two sizes, then it would require more clicks with just the increase and decrease buttons than it would with a drop-down menu. A drop-down menu for text sizes would conserve space, provide the user with more information, and reduce the number of clicks the user has to make in certain scenarios.

Another recommendation to make is to rename the “object” button to “insert image.” Audio and video files are not permitted, but can still be linked to in the messages [252]. Since only image files will be supported, it would be best to change the name of the “object” button to reflect that. Another problem with the name of the button is that a user could potentially mistake the word “object,” meaning “an item,” for its verb heteronym, which means “to disagree with something,” and believe that the button has an entirely different purpose. In order to prevent user confusion, renaming the button is highly recommended.

There are also modifications that can be made to the notes generated by the “refer” and “source” buttons. The original design for the conversation page does not explain what happens if text that has been highlighted by either the “refer” or “source” buttons is scrolled off-screen in the scrollable box. In the example image shown in the original document, the only notes displayed are ones linked to by text that is on-screen in the scrollable box [252]. As the effect of highlighted text being scrolled off-screen is unclear, it is recommended to ensure that notes generated by the “refer” and “source” buttons remain visible to the side of the scrollable box regardless of whether or not the text they are linked to is on-screen at the moment. This recommendation is made in conjunction with a recommendation that clicking on a note should have the scrollable box jump to the position of the text that the note is linked to, and should also emphasize the text it is linked to for increased ease of viewing. These recommendations will allow a user to re-examine the full context

of a note at will without having to manually scroll to the position of the highlighted text that the note is linked to. These recommendations will also remove the need to use arrows to signify which note is associated with which piece of highlighted text. However, if these recommendations are implemented, there comes the possibility of a scenario in which there are more notes created by the “refer” and “source” buttons than can be displayed on the conversation page. A possible solution to this problem would be to make the area to the side of the scrollable box into a second, separate scrollable box which contains notes created by the “refer” and “source” buttons. This would unfortunately cause the user to have to scroll through the note box in order to find the note they are looking for, but it would allow the user to be able to search through the contents of previous messages in the conversation by simply clicking on the associated note rather than scrolling through the previous messages in search of a particular note or section of text.

Comments made on documents created in Google Drive have a similar functionality to what is described in the previous paragraph. Comments in Google Drive are made by highlighting a portion of text in the document and using a menu command to insert a comment which can contain both text and links, similarly to how the “refer” and “source” commands are described in the original document [157, 252]. Text that has a comment linked to it becomes faintly highlighted, and clicking on either the comment or the text that the comment is linked to will result in the color of the highlighting changing to a deeper shade in order to differentiate one portion of highlighted text from another [157]. Additionally, documents in Google Drive have a button at the top of the page called “Comments,” which displays a scrollable box containing all comments present in the document [157]. Clicking on a comment in this box will automatically result in the document jumping to the location of the text that the comment is linked to, as well as emphasizing

the highlighted text associated with that comment [157]. The Civil Good conversation page could implement similar functions to those used in the Google Drive comment system in order to improve the functionality of the notes generated by the “refer” and “source” buttons.

Lastly, one characteristic of computer-mediated communication that has not been addressed by the current model is the silence in between each response (known as “dead air”). In face-to-face conversation, dead air is avoided and considered socially awkward [420]. In instant messaging, the time users spend typing out their responses can be interpreted as dead air, especially taking into account varying typing speeds between the users [420]. This can be eliminated with some indicator of user status (typing or reading) on the user interface [420]. It is therefore recommended to implement a user status indicator in the conversation interface. The majority of modern instant messaging applications (e.g. Skype or AIM) have some indicator of the conversation partner’s activity. For example, Microsoft’s Windows Live Messenger displays the text “someone is typing” while the user’s partner is typing a message [418]. For Civil Good conversations, a user status indicator would let the user know if their partner is online and still present in the conversation in order to help prevent an uncomfortable sense of dead air in between messages.

### **13.3.3 Conclusion**

There are a number of improvements that can be made to the conversation page in order to conserve space and prevent confusion. Seeing as the “undo” button does not provide any meaningful function, it should be removed. Similarly, the two buttons that change the text size should be consolidated into a single drop-down menu of text sizes. The “object” button should be renamed to “insert image” in order to avoid potential misunderstandings and to reflect the fact that only image

files will be supported on the conversation page. In order to reduce unnecessary scrolling and improve user ability to search through messages in a conversation, the area to the side of the scrollable box containing conversation messages should be made into a second scrollable box which contains notes generated by the “refer” and “source” buttons. Furthermore, clicking on a note should make the box containing conversation messages jump to the location of the text that the note is linked to and should emphasize said text for increased readability. If these recommendations are implemented, then the arrows that are used to signify which note corresponds to which piece of highlighted text will no longer be necessary and can therefore be removed as a feature. Additionally, as was recommended in Section 5.5.1, it would be useful to include a “report” button in order to flag inappropriate content or behavior or to label a conversation as containing content not suitable for minors. It is also recommended in Section 13.2.3 that the names of conversation participants not be shown during conversations. Therefore, there is no need for there to be a link to a conversation partner’s profile. The only exception would be in the case where two users who have Faved each other choose to engage each other in conversation, as Faves are naturally able to view each other’s display names. Lastly, the activity status of a user’s conversation partner should be displayed in order to reduce “dead air” during real-time conversations.

## **13.4 Rating System**

The conversation rating system is a major feature of Civil Good. It is used to determine the quality of a user’s conversational skills on the site and can be used to potentially match people up during the seek process. We thus researched the optimal rating system. Three key aspects were examined. They were the number of criteria used, the type of criteria to be used, and the format by which the criteria should be

rated. Based on the optimal rating system found, we then recommend changes to the original rating system proposed in the Civil Good design document [252].

### 13.4.1 Background

With regard to the number of criteria, studies have been done to analyze the use of multiple criteria in systems that recommend items to users. In general, the studies have concluded that the use of multiple criteria in these recommendation systems can help produce accurate and precise recommendations to users [302, 401, 12]. This supports using multiple criteria in a rating system that would recommend conversations to users. However, earlier research in Section 13.1.1 concludes that having more choices tends to have a negative impact on the willingness of users to participate. This suggests that a minimal number of criteria should be used in order to cover sufficient facets of the rating system but still keep users interested.

With regard to which criteria to use research suggests that Grice's Cooperative Principle could apply as a guideline for which criteria to use. In summary, this principle suggests that the average person assumes that conversation should adhere to four different maxims in order to achieve meaningful cooperation: quantity, quality, relation, and manner [170]. Quantity, is ensuring that enough information is given to significantly describe the speaker's intentions without going into excessive detail. Quality is when the provided information is true and can be backed up with evidence. Relation is making sure that the information given is relevant to the topic of conversation. Lastly, manner is ensuring information is conveyed clearly and understandably.

Two studies were found that suggest that the four maxims of the Gricean Cooperative Principle do help produce meaningful conversation. The first study was conducted by Chia-Huan Ho and Karen Swan on the students of an online graduate-



level English grammar class [182]. The study analyzed online posts made by students using the Gricean maxims and compared it to their course performance. The results of the study suggested there was a general positive relation between course performance and the degree that the students' posts followed the Gricean maxims. The second study analyzed academic research papers to devise a framework based on the Gricean Cooperative Principle for how to use metadiscourse (discussion on how converse or debate) in writing future research papers [7]. The study resulted in a tentative model, but during the process it was found that the Gricean Cooperative Principle can be assumed to have influence in the decisions made in all writing for communication.

There is one limitation in regards to these two studies: both of the studies were limited in scope. The student posts and research papers that were analyzed were from people of a higher academic standing. Not every user of a website will have the same background as these people. The rating system could potentially have different effectiveness on non-academic people.

Additionally, the Gricean maxims do not cover the element of politeness [417]. A person can still manage to be rude while following all four of Grice's maxims. Therefore, politeness should be added as a criteria in addition to all four of Grice's maxims.

Another problem for these criteria is the labels used. The maxims of quality and manner have modern day definitions that differ from Grice's intended meanings [269, 268]. Thus, to avoid ambiguity it is more optimal to find synonyms for the names of these criteria that match the intended meaning. Accuracy and clarity have unique definitions that are closer to Grice's intended meanings for quality and manner respectively [269, 268]. This suggests to use these as names for the criteria instead of quality and manner.

The last aspect of the rating system is the format used to rate the chosen criteria. E. Isaac Sparling and Shilad Sen conducted an experiment to analyze the use of four different rating scales [381]. The scales were unary (“like it,” such as on Facebook), binary (thumbs up/down, such as on YouTube), 5-star (such as on Amazon), and a 100-point slider. The results showed that the 5-star scale provided the greatest overall user satisfaction.

### 13.4.2 Discussion

The rating system given in the Civil Good original design document had twelve different criteria [252]. Ten of the criteria, as shown in Table 7, fall within the scope of criteria suggested in Section 13.4.1. The other two criteria, “Overall” and “Persuasiveness/Power of Argument” can be evaluated from the combination of the other criteria. Thus, changing the criteria used from the original twelve in the Civil Good design document to the five suggested by the background research would simplify the rating system while still preserving the general meaning of the original twelve. This simplification has a more positive impact on user willingness to participate in the rating system as supported by research in Section 13.1.1. Therefore, we recommend that the five criteria from Section 13.4.1 be used in place of the original twelve criteria.

Table 7: Rating Criteria

Research Criteria	Corresponding Criteria from Civil Good Paper
Quantity	Brevity
Accuracy (quality)	Accuracy of claimed knowledge, Accuracy of claimed Opinions/Facts, Sources cited
Relevance	n/a
Clarity (manner)	Language skills
Politeness	Civility, Timeliness of responses, Open-mindedness, Ability to create an enjoyable discussion, Listening skills

The research system in the Civil Good design document also suggested using a 100-point slider format for the rating system. However, the background research showed that among the unary, binary, 5-star, and 100-point slider scales, user satisfaction was greatest for a 5-star format and not a 100-point slider. For this reason, we recommend that the 100-point slider format from the Civil Good design document be changed to a 5-star format.

### **13.4.3 Conclusion**

In summary, we suggest to make two changes to the rating system proposed in the Civil Good design document. The first is to change the twelve criteria to five, namely quantity, accuracy, relevance, clarity, and politeness. The second is to change the format of the rating system from a 100-point slider format to a 5-star format.

## **13.5 Experiment**

In order to test our recommended changes to the Civil Good website, we created two versions of mockups for the website. One version of the mockup was based on the original design [252]. The other version was based on the recommended changes to the original design made in Section 13.2. We then designed and carried out a study to gather data on which of the two website versions was more usable.

### **13.5.1 Background**

The System Usability Scale (SUS) is an industry standard that can be used to measure the usability of a system such as a website [46]. To evaluate the usability, the following response questionnaire is used:

1. I think that I would like to use this system frequently.

2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Answers to the questions are provided in the format of a scale from one to five. Since this is an industry format, it would be beneficial to use this in the usability study for Civil Good.

An additional advantage to the SUS questionnaire is that questions are distributed between “positive” and “negative” questions. “Positive” questions are those that imply that the system is performing better when the answer is towards the high end of the scale, whereas “negative” questions are those that imply that the system is performing better when the answer is towards the low end of the scale. This ensures that a participant putting down the same rank for all questions would be easily found (and the data discarded) since a logical dead end would be reached.

In terms of sample size, studies have shown that observing a sample size of four or five participants allows for the discovery of 80% of the usability problems for a specific product [237, 118]. Another study recommended extending the sample

size to six or seven participants but does not recommend expending additional effort towards securing any more “since the value of additional subjects falls off exponentially” based on the study’s experimental data [291]. Thus, this number provides a good baseline for the number of participants needed to evaluate the usability of each of the two website mockups of Civil Good.

### **13.5.2 Methodology**

The goal of the usability study was to gather feedback to benchmark the user experience for the two separate versions of the website in order to determine which version of the website enabled a smoother, more enjoyable user experience. The two versions of the website were <http://civil-good.wpi.edu/>, which was the mockup based on the original design (the original version), and <http://civilgood.wpi.edu/>, which was the version based on the recommended changes (the modified version).

For this study, static frame mockups were created for each of the two versions of the Civil Good website. This means that HTML pages were created to display various features on the website but the pages had no functionality other than being able to navigate to other pages and fill out any forms.

In order to evaluate the mockups, a list of tasks to accomplish was created. These tasks involved navigating through the mockups to access various features that would be implemented in the actual Civil Good website. The list of tasks was the same for each version of the mockups.

These tasks were used in our usability study. For this study, students from Worcester Polytechnic Institute (WPI) were asked to participate. Participants were asked through email to come to the location of the study. They were met by a member of the group and the study was explained to them. After reviewing and signing the Informed Consent Agreement, the participants could begin the study.

First, participants of the study were asked to complete a Pre-Experiment Questionnaire. This questionnaire asked each participant for basic demographic information, the participant's self-assessed skill with computers, and the amount of time the participant generally spent on the Internet.

Next, the participants were handed the list of the tasks that they had to complete for the study, and were shown to a laptop which had the website homepage loaded. The version used by each participant was decided (and noted) beforehand, with care taken to ensure that an equal number of participants were testing either version. The participants were not informed that there were two separate versions of the website. The study was presented as a usability study on a website rather than a comparative study between two websites. This was done to avoid constraining the participant to just contrasting websites but instead encourage the participant to think openly on how to improve the website.

Once the participants had completed as many of the assigned tasks as possible, they were asked to fill out a Post-Experiment Questionnaire. The first part of this questionnaire contained the questions from the SUS as described in Section 13.5.1. This was used to quantitatively evaluate how well the participants were able to use the mockups to complete the tasks. The second part of the Post-Experiment Questionnaire contained additional questions that allowed users to provide qualitative feedback. At the end of this questionnaire, the study was concluded.

### **13.5.3 Discussion**

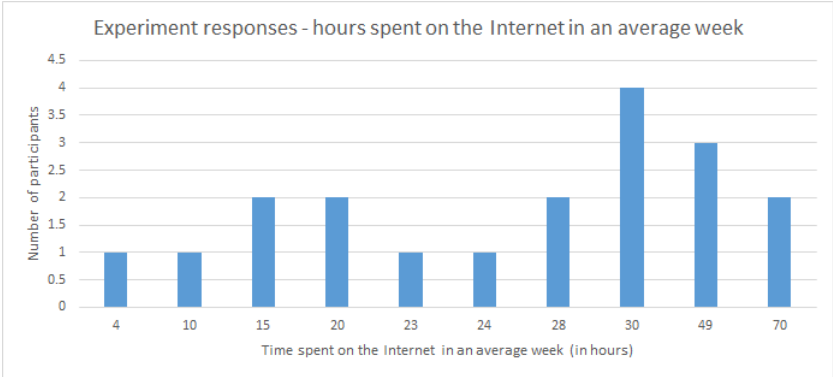
There were 20 participants who completed the usability study. The participants had a varying range of self-identified computer skill levels (on a scale of 1 to 5, with 1 meaning "Little" and 5 meaning "Advanced"), which are summarized in Table 8.

The distribution of (self-reported) time spent by the participants on the

Table 8: Experiment Data—Number of participants self-identifying at a particular level of computer skill

Skill Level	Number of participants who self-identified at that skill level
1	0
2	0
3	6
3.5	1
4	5
4.5	1
5	7

Figure 4: Experiment data—Hours spent on the Internet in an average week



*Note.* For this graph, only 19 responses out of 20 were considered. The response that was not considered was a participant stating that the time the participant spent on the Internet was “A lot”—since it was not numeric, it was not included in the graph.

Internet in an average week was also fairly skewed above the average (31.263 hours per week), as seen in Figure 4.

The skill level varied but was firmly in the above-average range of the reported data. This can be explained by the fact that the sample population was drawn from the student body of WPI, and university students would be fairly well-versed with computers. The distribution of time spent on the Internet can also be explained by the background of the sample population. While these distributions reduce the diversity of the population, they also imply that the sample has a higher level of

experience with computers and websites, which would indicate a correspondingly higher expectation of user interaction flow from a website. This higher expectation of a smooth user interaction would, naturally, generate feedback that would help improve the user experience to a greater degree.

The responses from the participants that carried out the tasks on the original version of the website are presented in Table 9.

The responses from the participants who completed the tasks on the modified version of the website are presented in Table 10.

To identify trends, we averaged the rankings of either website version for each question. The results of this calculation are presented in Table 11.

We see the following trends:

- Participants using the modified version would like to use the system more frequently than those using the original version.
- Participants using the original version found the system more unnecessarily complex compared to those using the modified version.
- Participants using the modified version found the system significantly easier to use compared to those using the original version.
- Participants using either of the two versions found that they would need the same level of help from a technical person to be able to use the system.
- Participants using the original version found the various website functions were integrated slightly better than those using the modified version.
- Participants using the modified version found less inconsistency in the system compared to the participants using the original version.



Table 9: Responses: Original Design

Participant	1	2	3	4	5	6	7	8	9	10
Skill level	5	5	3	3	4	4	5	5	5	3
Time on Internet	28	23	15	70	28	30	49	24	30	4
I think I would like to use this system frequently.	3	2	5	3	3	4	4	4	2	2
I found the system unnecessarily complex.	4	4	4	5	4	2	2	3	3	3
I thought the system was easy to use.	1	2	3	3	2	4	4	3	3	2
I think that I would need the support of a technical person to be able to use this system.	1	3	1	1	2	1	1	1	1	1
I found the various functions in this system were well integrated.	3	4	4	3	3	4	3	4	4	3
I thought there was too much inconsistency in this system.	2	5	4	4	3	2	2	2	2	2
I would imagine that most people would learn to use this system very quickly.	2	2	3	2	1	3	3	3	4	2
I found the system very cumbersome to use.	1	3	2	4	5	3	2	2	3	4
I felt very confident using the system.	2	2	4	3	2	4	3	4	3	1
I needed to learn a lot of things before I could get going with this system.	4	4	1	2	4	2	2	2	1	4

*Note.* 1 means “Strongly Disagree” and 5 means “Strongly Agree.”

Table 10: Responses: Modified Design

Participant	1	2	3	4	5	6	7	8	9	10
Skill level	5	4.5	4	3	3	4	3.5	3	4	5
Time on Internet	30	70	10	20	30	“A lot”	15	49	20	49
I think I would like to use this system frequently.	4	4.7	4	3	4	3	3	3	2	3
I found the system unnecessarily complex.	2	3	2	2	3	2	2	2	3	2
I thought the system was easy to use.	4	3.9	5	4	3	4	4	3	2	3
I think that I would need the support of a technical person to be able to use this system.	1	0	1	1	3	1	2	1	1	2
I found the various functions in this system were well integrated.	4	2	4	4	4	4	4	4	1	3
I thought there was too much inconsistency in this system.	1	1	2	2	3	1	3	3	4	2
I would imagine that most people would learn to use this system very quickly.	4	4.6	4	4	4	4	4	3	2	4
I found the system very cumbersome to use.	1	2.3	1	1	2	2	2	3	4	2
I felt very confident using the system.	5	4	5	3	4	4	3	3	2	4
I needed to learn a lot of things before I could get going with this system.	2	4.1	1	2	2	1	2	2	2	2

*Note.* 1 means “Strongly Disagree” and 5 means “Strongly Agree.”

Table 11: Experiment Data—Average ranking and standard deviations for each question for the two website versions

Question	Original		Modified	
	Avg	Stdev	Avg	Stdev
I think I would like to use this system frequently.	3.2	1.0328	3.37	0.7832
I found the system unnecessarily complex.	3.4	0.9661	2.3	0.483
I thought the system was easy to use.	2.7	0.9487	3.59	0.8386
I think that I would need the support of a technical person to be able to use this system.	1.3	0.6749	1.3	0.8233
I found the various functions in this system were well integrated.	3.5	0.527	3.4	1.075
I thought there was too much inconsistency in this system.	2.8	1.1353	2.2	1.0328
I would imagine that most people would learn to use this system very quickly.	2.5	0.8498	3.76	0.7291
I found the system very cumbersome to use.	2.9	1.1972	2.03	0.9476
I felt very confident using the system.	2.8	1.0328	3.7	0.9487
I needed to learn a lot of things before I could get going with this system.	2.6	1.2649	2.01	0.8439

*Note.* 1 means “Strongly Disagree” and 5 means “Strongly Agree.” “Avg” refers to the average of the data; “Stdev” refers to the standard deviation.

- Participants using the modified version agreed that most people would learn to use the system significantly quicker than the participants using the original version.
- Participants using the original version found it more cumbersome to use compared to the participants using the modified version.
- Participants using the modified version were significantly more confident when using the system compared to the participants using the original version.
- Participants using the original version thought that they had to contend with a higher learning curve before being able to really start using the system while participants using the modified version thought they were able to start using the system without much of a learning curve.

Using this sample population's results, we observe that the participants in the study who used the modified version of the website rated it higher on the questions regarding usability and the user interaction. The sole exception to this was the integration of the various functions, where the original version was rated to be slightly better. Both versions scored equally in terms of the participants thinking that they would need help from a technical person.

A sample size of 10 participants (per version of website) is more than the seven participants that other usability studies have found are enough to generate statistically significant results (as discussed in Section 13.5.1). Thus, the results found in the usability study are statistically significant enough to indicate that the modified version of the website promotes an increase in user interaction and general usability.

The comments provided by the participants using either version of the website also provided some points to consider for future design decisions. A number

of the participants indicated that the Conversations feature, involving starting a new Conversation, was difficult to find. For a website primarily revolving around fostering conversation between users, this is a significant piece of feedback. If the Conversations feature remains difficult to find on the actual Civil Good website, it could potentially frustrate users and detract from their motivation for using Civil Good.

It was also reported by the participants that the homepage of the website was cluttered—this was noted for both the original and modified versions, albeit with different frequencies. Clutter can decrease the readability of the homepage due to information overload, and this feedback demonstrates that the experiment participants preferred a more streamlined page.

In addition, the color scheme was reported by a few participants to be not pleasing to the eye—one participant in particular recommended increasing the contrast along with changing the colors. Incorporating feedback on colors and contrast can increase the readability of the content, and increase the rate of user satisfaction.

Participants were slightly confused by the practice of asking for two passwords on the signup page—it was actually observed that some participants interpreted the second password entry field as a ‘repeat password’ feature, instead of the second (separate) password feature that it is actually meant to be. As discussed in Section 9.4, two separate password actually decrease the security of the user’s account. In addition, after observing that users do not realize that the two passwords are meant to be different, we recommend that this feature should be removed.

Participants also reported that having a brief tutorial that explained a *Seek*, a *Fave*, and the different types of conversations like *Debate*, *Query*, *Devil’s Advocate*, *Whine*, *Gab*, and *Rant* would help decrease the learning curve associated with the system. A tutorial or help system that would be available to the users could help

decrease the learning curve associated with Civil Good.

Lastly, in terms of what was not mentioned by participants, the participants using the modified version of the website did not comment negatively on the presence of the textual advertisements at the bottom of the page. Thus, advertising could not affect the user experience if the advertisements were unobtrusive.

#### **13.5.4 Conclusion**

The data collected during the usability study indicates that the modified version of the website ranks higher in usability and user preference in a situation where the user is unaware that there is any other version of the website available. Thus, it indicates that the recommended changes to the website design succeed in improving usability in practice. We can thus make the conclusion based on the data that the recommended changes be incorporated into the design of the website when it is constructed.

In addition to those changes, we recommend that *Conversations* should be made a menu option. Clicking on the menu option should open a page with information about active conversations, unrated conversations, finished conversations, and the option to start a new conversation.

We also recommend that the amount of information contained on the homepage should be reduced and organized into sections better (for example, if the *Conversations* page was implemented, the entire section on *Finished Conversations* could be removed from the homepage).

Finally, we recommend that a brief tutorial be implemented for each new user. This tutorial could be as simple as a dynamic hover text box that opens for each page that the user is visiting for the first time.

## 14 Mobile Presence

With the proliferation of mobile devices, it is becoming increasingly important for every major service to offer a seamless mobile experience. We researched the possibility of ensuring a secure browsing experience on the major mobile platforms, Android and iOS [52], and provided a description of the differences between the website itself and the corresponding mobile applications.

### 14.1 Mobile Demographic

A website like Civil Good that will just be starting out needs as much exposure as it can get. One way to increase exposure is to market the website to mobile audiences. It is important to understand the mobile demographic in order to determine the best mobile option for Civil Good.

#### 14.1.1 Background

The American mobile demographic is quite substantial. As of May of 2013, 91% of Americans own a cell phone and 56% own a smartphone [45]. Additionally, 28% of cell phone users own an Android phone, 25% own an iPhone, and 4% own a Blackberry [45, 120]. Among mobile device-owners, 31% of males and 26% of females own an Android phone, while 24% of males and 26% of females own an iPhone [120]. Also, 24% of Americans of age sixteen and older own an e-Book reader, while 35% own a tablet computer [45].

Among the mobile demographic, a significant number use their devices for online activities. 63% of cell phone users use their devices to access the Internet, and 50% use them to download applications and add-ons [45, 255]. As of May of 2013, 34% of those who access the Internet using mobile devices use their devices

“mostly” for that purpose [255]. The majority of those who use their mobile devices primarily for Internet-related activities are from younger age brackets. 45% to 50% of mobile Internet users between the ages of 18 and 29 use their devices primarily to go online [255, 101]. By comparison, 35% of those aged 30 to 49, 14% of those aged 50 to 64, and 10% of those older than 65 use their mobile devices mainly for online use [255]. Regarding the younger demographic, 43% of mobile users between the ages of 18 and 24 own an Android phone compared to 31% owning an iPhone [120, 101].

When comparing mobile activities for ethnic demographics, certain trends can be identified. 60% of Hispanics were found to be “cell-mostly Internet users,” meaning that they access the Internet primarily through mobile devices. This number is significantly higher compared to 43% of black non-Hispanic Americans and 27% of white non-Hispanic Americans [255]. The data also suggests that lower education levels and lesser income coincide with higher rates of being cell-mostly Internet users. 45% of those with high school graduate-level education or lower use their mobile devices mainly to access the Internet, compared to 21% of those with a college degree [255]. Of those with college degrees or higher, roughly 38% own iPhones while 29% own Android phones. 45% of mobile users who earn less than \$30,000 per year use their devices mostly for Internet use, compared to 27% of those earning more than \$75,000 per year [255]. Of those earning more than \$75,000 per year, 40% own an iPhone while 31% own an Android phone [120].

The mobile demographic is a major component of many websites. In particular, in June of 2012, Facebook had over 543 million users who accessed the site through mobile devices. One hundred and two million of those users used only mobile devices to access the site [88]. In 2012, it was found that roughly 30% of new users in India registered for the site using mobile devices. Additionally, 90% of Facebook users in Africa use mobile devices [88]. Mobile users were also found to

be 20% more likely to use the site on a given day than desktop users [88].

Similarly, Twitter sees around 200 million users access the site through mobile devices at least once per month [400]. Mobile Twitter users are 79% more likely to use the site multiple times per day compared to the average user, and are 47% less likely to use the desktop version of the site [400]. Twitter users between the ages of 18 and 34 are also 21% more likely to use mobile devices to access the site [400].

#### **14.1.2 Discussion**

If Civil Good were to have an application or version of the website optimized for mobile devices, there would be a sizeable market for it. Furthermore, as shown by Facebook and Twitter, users of mobile devices tend to access those websites more frequently than the desktop versions of the sites. Considering that Civil Good is also a website that primarily deals with exchanging messages between users, it is therefore possible that users would access a mobile version of the Civil Good website more frequently than they would the desktop version.

Certain trends can be used to guide Civil Good to best serve the mobile demographic. For example, considering that those with high school-level education or lower are more likely to be “cell-mostly” Internet users, Civil Good could address that portion of the demographic by adding special focus to the information-gathering aspects of the site. Since 60% of Hispanics use their mobile devices primarily for online activities, it would be worthwhile to add a Spanish version of Civil Good, as well as multiple other languages, at some point in the future. Also, mobile users of social media sites such as Facebook and Twitter are more likely than the average user to access those sites multiple times per day. It would therefore be wise to ensure that Civil Good temporarily remembers users’ login sessions so that if they wish to re-access the site later in the day, they will not have to re-enter their information



each time. If a user's device is lost, then there is the possibility that their login session will end before someone else finds and picks up the phone. However, this would put the user's Civil Good information at risk in the event that the device is dropped and immediately picked up by somebody else, or if the device is stolen.

### **14.1.3 Conclusion**

Considering how massive the mobile demographic is, it would be in Civil Good's best interest to make a mobile version of the site. A large percentage of the global population use mobile devices to access the Internet, many of whom exclusively use mobile devices for online purposes. Sites like Facebook and Twitter receive a large percentage of their traffic from mobile device-owners, especially considering that mobile users are more likely than the average user to use those sites on a given day. By making a mobile version of Civil Good, the site would receive a great deal more exposure, which would help in its short-term and long-term growth.

## **14.2 Mobile Application Security**

The main issue when considering whether or not to make a Civil Good mobile application is security. A recent study determined that 48.4% of the over 225 thousand applications for iOS devices access a device's unique identifier, while 13.2% access its location, 6.2% access its address book, and 1.6% access the device's music library [363, 13]. While not to as great an extent as iOS applications, Android applications were also found to frequently access private user information, even information that is not relevant to the functionality of the application itself [13, 364]. With so many mobile applications capable of accessing private user information and sending that information to advertisers, it is important to determine whether the privacy of Civil Good users can be protected if a mobile application for the website

is produced.

### 14.2.1 Background

An application that has access to seemingly innocuous information such as a user's music or photo libraries, let alone more sensitive data such as the user's location, contact lists, or device unique identifier, can use that information to track a user's behavior without their knowledge [13]. Additionally, applications that connect to the Internet to perform their functions can use that connection to secretly transmit user data to advertising servers. More importantly, they often do not encrypt the user's data before sending it, resulting in a major risk of such data being hijacked and read by another source [310]. The accessing of private data does not even have to be without the user's knowledge. Developers for Android mobile applications are obligated to notify users what data a particular application will attempt to access, and the user must agree to allow the application that access in order to utilize said application [184, 310]. However, developers are not required to inform users as to what they plan to do with that particular data or how securely that data will be handled. Users may potentially expose large and unnecessary amounts of personal data to risk just to use an application [310].

One of the many risks of using a mobile device is that in order communicate from nearly any location, it is almost always seeking connections. Malicious programs can easily exploit a device's need for connection [230]. Mobile malware programs can target a device by attempting to disguise themselves as applications or add-ons. When a user unknowingly allows one to install itself, the program then gains full access to the device. This is because antivirus and antispyware programs are not nearly as common in mobile devices as they are in regular computers [227]. These mobile malware programs can steal user data or force the device to send calls

or text messages to raise the device-owner’s phone bill. If the device uses Bluetooth, the malware can also spread to other devices through Bluetooth signals [230, 227]. The constant connection that mobile devices try to maintain makes them highly vulnerable to malicious programs that attempt to abuse that connection.

Despite the numerous risks, it is possible to protect the privacy of mobile application users. ProtectMyPrivacy is an iOS application developed by University of California researchers that shows users what data their other applications are trying to access and offers recommendations to improve users’ privacy settings [363, 13]. The ProtectMyPrivacy (“PMP”) application alerts the user when one or more applications attempts to access personal data on the user’s device, and allows the user to allow or block that access selectively based on the user’s own discretion [363]. At the same time, if the other application’s access is denied by the user, the PMP application feeds the application in question anonymous fake data in place of the user’s own private data [13]. PMP is able to give suggestions as to which applications should be allowed or denied access based on voluntary feedback received from users of the application. As such, PMP has recommendations for 97% of the 10,000 most popular iOS applications, which more than two thirds of the application’s user base choose to follow [363].

Android users are a slightly different case. As mentioned before, in Android’s Google Play store, it is mandatory for application designers to specify any and all data that a given application may decide to access. As such, it is up to the user to either agree to allow the application access to that data or forego using that application entirely [184]. Users must choose between the functionality of a given application and the preservation of their private data. However, there is a way around this ultimatum. AppFence is an application that lets users block other applications from accessing data that is not necessary for them to perform their

advertised function. Additionally, if such private information is actually necessary for the application to access in order to function, AppFence can prevent the application from transmitting that data to advertisers. Similar to how PMP functions, AppFence provides false data in place of private user data when the application requests it [184].

There are several ways that mobile users can improve the security of their devices. Above all, a strong password and strong policies for password protection are a must in the event that the device itself is stolen [175]. If a device does wind up being lost or stolen, there are ways to remotely wipe the data from the device so that if somebody else finds the device then that person will not be able to access the owner's information [123]. Having a firewall and antivirus protection on mobile devices is highly recommended [175, 123]. Lastly, software similar to PMP and AppFence that can encrypt either device contents or communications can also help to improve device security [175, 123].

### **14.2.2 Discussion**

As described in Section 14.2.1, there are numerous threats to the security of mobile device-owners. Some come from mobile malware, others come from the various applications people use on a day-to-day basis that access user data. In order to mitigate this risk, it is recommended that Civil Good store minimal information on a user's device. A simple method to ensure that vital information is not accessed by applications on mobile devices is to not store vital information on mobile devices in the first place.

Additionally, it is recommended that Civil Good inform its users to take proper precautions when using mobile devices. Protective applications such as ProtectMyPrivacy and AppFence can improve mobile users' ability to prevent other

applications from accessing protected personal data without permission. Similar safe browsing habits such as choosing a strong password, maintaining a capable fire-wall and antivirus software, encrypting connections, and avoiding unsafe networks should also be recommended to Civil Good users.

### **14.2.3 Conclusion**

Mobile devices face numerous malicious entities in the form of devious applications and malware programs that seek to access private user data without permission. However, Civil Good is capable of making recommendations to mobile users that can help them reduce their likelihood of falling victim to these entities. Furthermore, storing minimal data on user devices will mitigate the danger of a user's Civil Good data being compromised by other applications or programs.

## **14.3 Mobile Web Design**

A website designed for mobile devices will have a very different layout compared to a desktop-based website. Therefore, it is important to understand the differences in design and functionality between a mobile website and a desktop-based one in order to evaluate the possibility of making a Civil Good mobile website or mobile app.

### **14.3.1 Background**

Since mobile devices have smaller screens than desktops and laptops, unnecessary content such as superfluous graphics and wasted space must be removed while crucial site features are promoted and made more easily readable. In addition to this, mobile websites are structured vertically as opposed to horizontally so that each topic or link has its own row, reducing clutter and making them more easily viewed and interacted with [247, 374]. Cutting away all but the bare essentials from a desktop-

based website design allows mobile websites to make topics larger and more spaced out so that users do not have to struggle to click on a link while using a device with a small screen [374]. Similarly, mobile websites also tend to use little to no hypertext, in order to prevent users from accidentally activating links with their fingers when they meant to scroll the page down [247].

A major structural difference between mobile and desktop based websites is that mobile websites have greatly reduced content organization hierarchies [247]. This means that users do not need to navigate through numerous categories sub-sections in order to reach a particular page. As such, “breadcrumbs” (a series of links showing which subcategories have been traversed in order to reach the current page) and contextual navigation are not needed for mobile websites, and the space they would occupy can instead be used to prioritize content, rather than navigation links [247]. Aside from providing extra space by removing breadcrumbs, mobile website pages have less of a content hierarchy because their browsers are severely limited by bandwidth. This limited bandwidth restricts their ability to access and transfer data from a Web server. Mobile website developers want there to be as little hierarchy in a website’s pages as possible so that they can minimize the number of client requests and the amount of data transfer necessary for the user’s browsing experience [232].

Mobile websites are also capable of using a mobile device’s functions and data in order to provide more personalized options for the user. For example, mobile retail websites can allow a user to directly call or text a particular business in order to place an order, or they can access the mobile device’s location in order to find businesses near the user [247, 139]. Mobile real estate websites can provide users with price ranges and other details about various properties near them, and arts and entertainment websites can alert users to upcoming local events, all without

requiring any input from the user [139]. Local governments can even use mobile technology to send out alerts or provide the community with polls and surveys [139].

### **14.3.2 Discussion**

Mobile websites should above all be user-friendly. Mobile websites remove unnecessary content and space from their pages and leave only what will be useful and easily accessible to the user. The reduced content hierarchy minimizes the amount of page-loading necessary, and the integration with a mobile device's functions and data allows for a more personal experience.

### **14.3.3 Conclusion**

While it would be possible to make a mobile version of Civil Good, such a site would have to be designed differently from the desktop version in order to provide a more user-friendly experience. The proposed Civil Good website requires a large amount of clicking between pages and text entry, two things that severely harm the mobile experience. Streamlining many of Civil Good's processes is of vital importance to creating a desirable mobile website.

## **14.4 Mobile Applications vs. Websites**

It is important to determine whether Civil Good should engage the mobile demographic using either a website optimized for mobile devices or a completely separate downloadable application.

### **14.4.1 Background**

When deciding whether to make a mobile website or a mobile application, four important factors to consider are the purpose of the site, the audience of such a

site, the potential user experience, and the budget allotted for such a venture [215]. Developers must determine the functional requirements for mobile access before choosing between a website and a native application. It is also important to consider who would use a mobile version of a certain website, along with the purpose of the usage.

While mobile websites and mobile applications are both related to mobile devices, the two are in fact very different from one another in terms of design and functionality. For starters, applications tend to have superior graphics, effects, and performance overall, considering that a mobile website is limited by the device-owner's bandwidth while an application is only limited by the specifications and performance of the device itself [215]. Additionally, while applications are more expensive to produce compared to mobile websites, applications can easily be monetized by making users purchase them in an application store, as well as through the use of advertisements, whereas it is highly impractical to try to make users buy the ability to access a mobile website [215]. Mobile websites have the advantage over mobile applications when it comes to the ability to get overall exposure. A mobile website can be developed just as easily as a desktop-based one, whereas an application requires a completely different set of tools and skills [215]. Mobile websites can also be published immediately and are viewable instantly once published. While the same is true for Android applications, iPhone applications must go through an approval process just to enter a device's application store, and even then they must rely on a potential user discovering them in the depths of a device's application store [215]. On top of that, an application must be developed for each mobile platform individually, whereas a mobile website will show up and function on all types of mobile devices so long as they have an up-to-date Internet browser [215]. Lastly, mobile websites are easily updated and any changes take effect immediately, whereas



an updated version of an application must go through the same submission process, and even then the user must download and install the update for the changes to take effect [215].

When considering how a website should approach the mobile market, it is important to know how people use their mobile devices. People tend to use mobile devices for three things: information-seeking, communication, and content object handling [83]. Information-seeking refers to searching for information, from specific facts to general Internet browsing. Communication refers to email, social networking, and other online chat sites. Content object handling refers to downloadable objects such as ringtones, wallpapers and other add-ons, or personal content such as photos or videos. Regarding communication, user studies show that while mobile users use their devices to check email frequently, they rarely respond to emails, instead preferring to do so from a computer [83, 416]. A study from 2008 by the Nokia Research Center shows that mobile users reply to emails on their devices roughly 5% of the time [83], while more recently, around 9% of surveyed iPad users use their devices to respond to email [416].

#### **14.4.2 Discussion**

The first question to consider is what a mobile version of the Civil Good website would do. Civil Good can be used in order for people to obtain information, whether it be from conversations either current or archived, or from various polls and survey data. The site can also help people to learn about different arguments or opinions regarding various issues. Additionally, Civil Good can be used as a tool to help someone gauge the political climate or public opinion, both in the present and in the past. Lastly, Civil Good hopes to foster an environment that can mitigate bias, prejudice, and fear of having an unpopular opinion through the use of promoting

self-affirmation and reducing polarization and toxic disinhibition. Since Civil Good is a text- and conversation-based website, a mobile version of the site would likely provide essentially the same features as the regular website version, only to a mobile audience. Mobile applications have direct access to device hardware such as the camera, microphone and accelerometer; however, these features are unlikely to be used with a Civil Good mobile application and as such do not offer a real advantage. A feature unique to mobile applications is the ability to integrate with the device notification system. For example, the mobile application could send the user a notification whenever a conversation partner replies to a discussion post or sends a Fave request.

The potential audience for a mobile version of Civil Good would be much larger than the normal desktop audience. Websites such as Facebook and Twitter receive hundreds of thousands of unique mobile visitors per month, many of whom access said websites exclusively via mobile devices [88, 400]. Mobile users are also far more likely to access those websites multiple times per day [88, 400]. If Civil Good is made compatible with mobile platforms, then it has the potential to receive a large amount of traffic from mobile Web-users. As a text-based website, a mobile version of Civil Good can find appeal with mobile users that utilize similar text-based websites and applications such as texting services, digital reading programs, fact-checking sites, and debate sites. In addition, the text-based nature of Civil Good means that the site will load quicker than sites that utilize other media forms such as images and videos.

Regarding the potential user experience, Civil Good mostly deals with delivering information and exchanging messages between users. A mobile version of the site should be more streamlined. For example, in order to reduce hierarchy, notifications can appear through drop-down or pop-up menus on the homepage as

opposed to having to create and load a separate page for them. Since most mobile users check their email on their devices but rarely type out responses, it is likely that users will use the mobile version of Civil Good to check messages and browse open or archived conversations as opposed to creating seeks or typing out replies.

Because the process for developing mobile websites is similar to that of a desktop website, mobile websites are less expensive to develop than mobile applications. For example, the desktop website could employ responsive Web design, which means that the website would contain the content of both the mobile and desktop versions, but the site would automatically detect whether the user is on a mobile device or not and change the display style to suit the device [254]. This means that the mobile and desktop versions of the site could be developed at the same time, reducing cost significantly. Mobile websites are also less expensive because mobile applications need to be developed individually for each platform. However, although applications are more expensive to develop, they have the advantage of being able to be monetized by having users to purchase them in their devices' application stores. Despite this, it would be better to save money by developing a mobile website instead of a mobile application.

#### **14.4.3 Conclusion**

In conclusion, a mobile website would be more effective than a mobile application for the purposes of Civil Good. Since the site mostly focuses on pairing up users for conversation and exchanging messages between them, the superior graphical quality and performance of an application would be wasted on Civil Good. Furthermore, the ability of an application to function while offline would also serve no purpose for Civil Good since the entire focus of the site is the ability to exchange messages with other users—something that requires a connection between users.

For a website like Civil Good that will just be starting out, gaining exposure is perhaps its top priority in order to build up a user base and establish a community. While having an application as opposed to a mobile website would be a better source of revenue, the only way to find an application is to either search for it in a device's application store or to be linked to it from a website. This means that the main demographic that would use a mobile application for Civil Good is people who already know about Civil Good enough to search for an application for it themselves. In contrast, a mobile website can easily be found through standard searches or pop up in related searches on a mobile browser as easily as the regular website would on a desktop or laptop computer.

## 15 Conclusions

Based on our research, we make the following recommendations for the Civil Good concept:

- Civil Good should build a community that makes negative interaction socially unacceptable by inviting well-mannered individuals to be the initial members of Civil Good and encouraging them to participate in positive discussion, as discussed in Section 4.1.
- As recommended in Section 4.2, conversations between users should be limited to small groups, and preferably to a one-on-one format. Using one-on-one debate also allows Civil Good to prevent majority groups from silencing those with an unpopular opinion, as discussed in Section 6.2.
- As discussed in Section 4.3, Civil Good could utilize the following methods of self-affirmation:
  - Positive feedback (e.g. congratulatory messages) as a source of self-affirmation for the users.
  - Build a Civil Good user profile for each individual user as a place for users to review their accomplishments on the site.
  - Manage a non-debate oriented conversation system to allow users to talk about their own personal hobbies or values without forcing a “values essay” on the user.
- As discussed in Section 4.3, Civil Good should not promote awareness of the positive effects of self-affirmation.

- Civil Good could educate its users on bias-inducing psychological effects, as recommended in Section 4.4. For users that are already aware of these effects, Civil Good should only provide them with light reminders about their potential biases and allow them to adjust on their own.
- As discussed in Section 5.1, Civil Good should treat all user information as PII, even if it is not legally considered PII.
- Also as discussed in Section 5.1, Civil Good must continuously update the privacy policy, specifying what information will be collected, how it will be stored and protected, and what Civil Good will do in the event of a security breach.
- As elucidated in Section 5.1, Civil Good must post a policy explaining how the website responds to “do not track” signals.
- As discussed in Section 5.1, Civil Good should notify all users in the event of a security breach, even if this is not legally required.
- As discussed in Section 5.1, Civil Good should comply with the Federal Trade Commission’s “Fair Information Principles” and post its compliance with these guidelines.
- The DMCA requires Civil Good to develop a “notice and takedown” procedure for copyright infringement. Civil Good will need to respond to all notices in a timely manner, as explained in Section 5.2.
- As discussed in Section 5.3, Civil Good should comply with Section 230 of the Communications Decency Act by never altering the meaning of any potentially defamatory content posted by a user.

- As discussed in Section 5.4, if Civil Good receives a subpoena, legal counsel will need to be sought to determine whether Civil Good should comply with or fight the subpoena.
- As discussed in Section 5.4, if Civil Good receives an informal request for information from a government agency, Civil Good will have to decide whether or not to comply from a business and ethics standpoint.
- As discussed in Section 5.5, Civil Good should ask users for their age upon sign-up, so that policies involving minors can be properly enforced.
- As discussed in Section 5.5, children under the age of 13 should not be allowed to use Civil Good, in order to not be required to comply with COPPA.
- Also as discussed in Section 5.5, users under the age of 18 should be prevented from seeing inappropriate content.
- As discussed in Section 5.6, Civil Good should retain legal counsel so as to keep apprised of any new legal developments.
- Also as discussed in Section 5.6, Civil Good should clearly designate who is authorized to speak on behalf of the website, and implement policies that prevent other workers from claiming this authority.
- As discussed in Section 6.1, Civil Good should ensure its users are well-educated in the subject matter to prevent politically extreme views from causing polarization among its users.
- As discussed in Section 6.2, Civil Good should maintain strict privacy for its users to allow them to convey unpopular opinions without being attacked for their opinions.

- As discussed in Section 6.3, Civil Good could provide longitudinal data on political climate through analysis of its user conversations.
- As discussed in Section 6.4, Civil Good should consider advertising, community building, and monetization through a conference or other recurring live event.
- As discussed in Section 7.1, Civil Good should initially use Virtual Private Servers (VPS) for its web hosting and switch to a dedicated server if website traffic becomes too much for VPS to handle.
- As discussed in Section 7.2, a relational database management system is the recommended database management system for Civil Good.
- As discussed in Section 7.3.1, PostgreSQL and MySQL are equally viable database engines.
- As discussed in Section 7.3.2, Nginx is recommended over Apache for web servers.
- As discussed in Section 7.3.3, it is recommended that Civil Good only utilize HTML, CSS and JavaScript for its client-side languages.
- As discussed in Section 7.3.4, either the Python web framework Django or the Ruby web framework Ruby on Rails would be ideal for the development of Civil Good.
- As discussed in 7.3.5, it is worth looking into additional miscellaneous open-source technologies such as Aptana Studio and Twitter Bootstrap to aid in the development of Civil Good.
- As discussed in Section 8.3, incremental refinements are the recommended method of making changes to Civil Good once the site is already available to



the general public. It is recommended that development decisions be data-driven.

- As discussed in Section 9.1, Civil Good should run the database management system and website on separate servers.
- As discussed in Section 9.1, Civil Good should use public key encryption (SSL/TLS) for communication between the database and website servers.
- As discussed in Section 9.1, Civil Good should restrict database permissions to the minimum necessary for the web application.
- As discussed in Section 9.1, Civil Good should avoid SQL injection by sanitizing inputs before inserting data into the database.
- As discussed in Section 9.1, Civil Good should utilize software such as AMNESIA to mitigate the risk of SQL injection.
- As discussed in Section 9.1, Civil Good should avoid dynamically constructing and executing query strings using user input.
- As discussed in Section 9.1, Civil Good should use input validation to verify data integrity.
- As discussed in Section 9.1, Civil Good should avoid printing SQL error messages.
- As discussed in Section 9.2, Civil Good should sanitize any user-submitted data on output to prevent XSS attacks.
- As discussed in Section 9.2, Civil Good should not sanitize user input for XSS attacks; but instead only sanitize on input for SQL injection attacks.

- As discussed in Section 9.2, Civil Good should use HTTPS for all connections.
- As discussed in Section 9.2, Civil Good should use HTTP referer-watching for the login form.
- As discussed in Section 9.2, Civil Good should use strong CSRF tokens to validate requests from logged in users.
- During development, ensure that all code is tested and reviewed, as recommended in Section 9.2.
- As discussed in Section 9.2, Civil Good should rate limit connections to mitigate the effects of DDoS attacks.
- As discussed in Section 9.2, Civil Good should utilize an intrusion detection/prevention system such as SNORT to reduce the risk of unauthorized access and DoS attacks.
- As discussed in Section 9.3, Civil Good should use HTTPS to mitigate the risk of man-in-the-middle attacks.
- As discussed in Section 9.3, Civil Good should encrypt data in the database using symmetric key encryption.
- As discussed in Section 9.3, Civil Good should use encryption algorithms with long keys such as AES for encrypting stored data.
- As discussed in Section 9.3, Civil Good should keep encryption keys in a separate location.
- As discussed in Section 9.3, Civil Good should implement end-to-end encryption between the client and server, and should use asymmetric, public key encryption such as SSL/TLS.

- As discussed in Section 9.3, Civil Good should encrypt data such that the information's protection is mathematically provable to as close to perfect secrecy as possible.
- As discussed in Section 9.4, Civil Good should require only one password instead of two. As discussed in Section 13.5, the experimental data also suggested that the practice of asking for two separate passwords during the sign-up process should be removed.
- As discussed in Section 9.4, email addresses should be made optional so that users who want to provide an email address for password recovery have the opportunity, and users who desire complete anonymity are not compelled to provide an address. Any provided addresses should be encrypted.
- As discussed in Section 9.4, Civil Good's password policy should not have character restrictions, and should not require users to periodically change passwords. The only permissible restriction is minimum length.
- As discussed in Section 9.4, Civil Good should encourage users to use strong passwords.
- As discussed in Section 9.4, Civil Good should hash passwords with a proven hashing algorithm with a long output hash. Do not create a unique hashing function for Civil Good.
- As discussed in Section 9.4, password hashes should be salted with unique salts that incorporate data from both the application server and the database server.
- As discussed in Section 10.1, Civil Good should track the following pieces of information

- Civil Good should keep track of user sessions through cookies, HTTP authentication if cookies fail, and IP addresses if HTTP authentication fails.
- Civil Good should gather usability data through click stream analysis
- Civil Good should gather web analytics through server logs
- As discussed in Section 10.3, Civil Good should incorporate the following recommendations with regards to third party privacy risks
  - Civil Good either (1) allows advertisers to use cookies and web bugs but not reveal any more information about users to the advertisers or (2) allows advertisers to target a specific demographic but disallows cookies and web bugs at the disadvantage of not being able to prove to advertisers that the advertisements are effective.
  - Civil Good limits the code a third party can execute to prevent browser fingerprinting.
  - Civil Good adds to its privacy policy that it is not responsible for the user's IP address and therefore geographic location being determined by a third party if the user clicks on an advertisement.
- As discussed in Section 10.2, the following recommendations should be incorporated with regards to browser fingerprinting and the associated privacy risks
  - Civil Good should develop policies such that no information that can be used for browser fingerprinting is logged or stored together.
  - Civil Good should log IP addresses for a period of time in case they are needed to investigate an attack.

- As discussed in Section 10.4, Civil Good should develop a privacy policy that addresses concerns that users would have, and should enforce that privacy policy with the utmost strictness. Providing and adhering to a privacy policy increases the likelihood that users will feel safer when using the website. This can also help with user retention.
- As discussed in Section 10.4, the privacy policy should contain as little an amount of legalese as possible, and should be presented in a format that is easy to comprehend. We suggest using a bulleted list.
- As discussed in Section 10.4, the privacy policy should be easily accessible by users. We suggest providing a link to the privacy policy in the footer section of the website.
- A proposed Privacy Policy is provided in Appendix E. The privacy policy will have to be reviewed and finalized by Civil Good’s legal counsel before the website goes live. The privacy policy will have to be amended by Civil Good’s legal counsel in the future, as and when required by legal changes or changes to Civil Good’s policy, as elucidated in Section 10.4.
- As discussed in Section 11.1, Civil Good should consider open source development.
- As discussed in Section 11.1, Civil Good should provide an initial codebase as proof of viability and as a starting point for its potential developers.
- As discussed in Section 11.1, Civil Good should leave complex, non-menial tasks as a challenge and as an attraction to potential developers.
- As discussed in Section 11.1, Civil Good, as a concept, should attempt to appeal to the developer audience.

- As discussed in Section 11.2, Civil Good’s core audience should be those who are politically active: people of higher status, higher education, and who are involved with their communities.
- As discussed in Section 11.2, Civil Good should provide a low level of entry by providing conversation topics that are more accessible to the general public.
- As discussed in Section 11.2, Civil Good should balance the registration process to prevent new users from being turned away, yet discourage users from making throwaway accounts.
- As discussed in Section 11.2, Civil Good should establish a set of rules to build a sense of community.
- As discussed in Section 11.2, Civil Good’s developers should actively participate with the community.
- As discussed in Section 11.2, Civil Good should consider a form of account memorization to increase logged-in user activity.
- As discussed in Section 11.2, Civil Good should care for its aesthetic design to improve user perception of the site.
- As discussed in Section 12.1, Civil Good needs to consider the following costs:
  - about \$34.47 per month for web hosting
  - about \$11.63 per year for a domain name
- As discussed in Section 12.2, after Civil Good has become established, it should apply to become a 501(c)(3) non-profit organization.

- As discussed in Section 12.3, it is recommended that Civil Good pursue search engine marketing and email marketing but not display advertising.
- As discussed in Section 12.4, Civil Good should use targeted display advertising with platforms such as Project Wonderful and Adzerk, which give control over what advertisements are displayed.
- As discussed in Section 12.4, Civil Good should utilize text-based advertisements or skyscraper banners that blend with the site design.
- If desired, avoid including advertisements in the conversation web pages so as not to distract users, as discussed in Section 12.4.
- As discussed in Section 12.4, Civil Good should not solicit donations more than six times a year.
- As discussed in Section 12.4, Civil Good should utilize demographic information such as age and gender to determine the frequency with which to ask users for donations.
- As recommended in Section 12.4, an online store should only be started once Civil Good has built a large, loyal community.
- As recommended in Section 13.2, the following changes should be incorporated into the various user interaction scenarios
  - The signup process should only ask users for the absolute minimum amount of information. The user’s facts and opinions should be asked for during the process of constructing a Seek.
  - Polls should not be mandatory.
  - Conversations should be completely anonymous.

- There should only be one type of Fave.
  - Initially, users should be able to see every Seek that they are a match for. As the number of users and Seeks grows, the number of Seeks a user can see should be limited.
  - If a Seek’s criteria are too narrow, they should be automatically relaxed, and the user should be notified about which criteria were relaxed.
  - The choice of interactive or non-interactive conversations should be eliminated, and all conversations should be interactive.
  - The Instant Match option should be eliminated.
  - Civil Good should use gamification to reward users for contributing.
  - The website should be designed to be accessible to users with disabilities
- As discussed in Section 13.3, the following recommendations should be incorporated into the conversation page
    - The “undo” button is unnecessary from a logistical standpoint and should be removed.
    - The two buttons that increase and decrease the text size should be replaced with a drop-down menu that clearly defines the range of possible text sizes.
    - Since only image files are supported by the conversation page, the “object” button should be renamed to “insert image.”
    - Notes generated by the “refer” and “source” buttons should be displayed in a second scrollable box next to the one where conversation messages are displayed. Clicking on a note in this box should make the box containing conversation messages jump to the location of the text that note



- is linked to and should emphasize said text for increased readability. As a result, the arrows used to signify which note corresponds to which piece of highlighted text will no longer be necessary and should be removed.
- As discussed in Section 5.5.1, a “report” button should be added to the conversation page to either to flag inappropriate content or to mark a conversation as unsuitable to be viewed by minors.
  - As discussed in Section 13.2.3, the names and profile page links of conversation participants should not be shown during conversations, with the exception of Faved users engaging in conversation with each other.
  - A user activity indicator should also be implemented in order to reduce “dead air.”
- As discussed in Section 13.4, the following changes should be incorporated to the conversation rating system
    - Change the twelve rating criteria given in the Civil Good design document to quantity, accuracy, relevance, clarity, and politeness.
    - Change the format of the rating system given in the Civil Good design document from a 100-point slider format to a 5-star format.
  - As discussed in Section 13.5, we make the following recommendations from the usability study
    - As per the experimental data, incorporating the recommendations made in Section 13.2 increases both the usability and the overall consistency of the website. The changes to the website also help feel users feel more confident when using it, and decrease the learning curve associated with the website.

- The design should be changed such that the Conversations feature is easy to find on the actual Civil Good website. One easy manner of accomplishing this is to make the Conversations feature an option on the website’s header navigational menu.
  - The amount of information on the homepage should be reduced. This can be accomplished by organizing sections better under the various menu options.
  - The colors used for the website should be such that there is a high contrast between the background and the text, which could increase user satisfaction by increasing the content’s readability.
  - A tutorial or help system that explains the various features of the website would help to further decrease the learning curve associated with using the website.
  - Unobtrusive, text-based advertisements were not found to detract from the quality of the user experience.
- As discussed in Section 14.1, in order for Civil Good to gain exposure, it is recommended that a version of the site be developed for mobile platforms
  - As discussed in Section 14.2, Civil Good should keep minimal information on a user’s device in order to minimize security risks.
  - Also as discussed in Section 14.2, Civil Good should recommend that its mobile version users to take proper security precautions when using mobile devices.
  - As discussed in Section 14.3, a mobile version of the Civil Good website should be streamlined and have reduced content hierarchies in order to improve mobile user-friendliness.

- As discussed in Section 14.4, it would be better for Civil Good to focus on developing a mobile version of its website than a separate mobile application.

Areas in psychology, marketing, costs, monetization, law, and mobile presence were researched extensively to support these recommendations. These recommendations were made with the intention of helping Civil Good provide an environment where people can engage in deeper online discussion. While discussion and communication are regarded as the Internet's most prominent achievements, there is a distinct lack of an online service that allows people to discuss their opinions freely and openly without the threat of persecution. With the Internet's unique quality of being able to connect people across geographic boundaries and social barriers, there is an opportunity for a service that allows true freedom of speech to spark civil discourse on a wider societal scale and help people expand their worldview beyond their local setting. Civil Good looks to accomplish this with a service that leverages both anonymity and the capabilities of the Internet. With these considerations in mind, Civil Good could become an authority in intelligent discourse and social change.

## 16 Bibliography

- [1] Smith v. Maryland. *US Supreme Court*, 442(78-5374):735, 1979.
- [2] US v. Forrester. *US Court of Appeals, 9th Circuit*, 512(05-50410):500, 2007.
- [3] 41st Parameter. Complex device recognition.  
<http://www.the41st.com/solutions/complex-device-recognition>,  
2013.
- [4] 501c3GO. 501c3 advantages.  
<http://attorneys.501c3go.com/about-501c3/501c3-advantages/>, 2013.
- [5] 501c3GO. Attorneys & CPAs. <http://attorneys.501c3go.com/>, 2013.
- [6] Jesse Aaron. The complete guide to advertising on reddit. *Maximize Social Business*, October 2013.
- [7] Reza Abdi, Manoochehr Tavangar Rizi, and Mansoor Tavakoli. The cooperative principle in discourse communities and genres: A framework for the use of metadiscourse. *Journal of Pragmatics*, 42(6):1669–1679, 2010.
- [8] Alan I Abramowitz and Kyle L Saunders. Is polarization a myth. *Journal of Politics*, 70(2):542–555, 2008.
- [9] Dominic Abrams, Margaret Wetherell, Sandra Cochrane, Michael A. Hogg, and John C. Turner. Knowing what to think by knowing who you are: Self-categorization and the nature of norm-formation, conformity and group polarization. *British Journal of Social Psychology*, 29(2):97–119, 1990.

- [10] Spencer Ackerman. Fisa court order that allowed NSA surveillance is revealed for first time.  
<http://www.theguardian.com/world/2013/nov/19/court-order-that-allowed-nsa-surveillance-is-revealed-for-first-time>, 2013.
- [11] ACM Conference on Human Factors in Computing Systems. *Who are the crowdworkers?: shifting demographics in mechanical turk*, Atlanta, Georgia, April 2010. Association for Computing Machinery.
- [12] Gediminas Adomavicius and YoungOk Kwon. New recommendation techniques for multicriteria rating systems. volume 22, pages 48–55. IEEE, 2007.
- [13] Yuvraj Agarwal and Malcolm Hall. ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing.  
[http://www.cs.cmu.edu/~yuvraja/docs/Agarwal\\_MobiSys2013\\_ProtectMyPrivacy.pdf](http://www.cs.cmu.edu/~yuvraja/docs/Agarwal_MobiSys2013_ProtectMyPrivacy.pdf), 2013.
- [14] Mary Cindy Ah Kioon, Zhao Shun Wang, and Shubra Deb Das. Security analysis of MD5 algorithm in password storage. *Applied Mechanics and Materials*, 347:2706–2711, 2013.
- [15] Shaaron Ainsworth, Giulia Gelmini-Hornsby, Kate Threapleton, Charles Crook, Claire OMalley, and Marie Buda. Anonymity in classroom voting and debating. *Learning and Instruction*, 21(3):365 – 378, 2011.
- [16] Alexa.com. How popular is ted.com?  
<http://www.alexa.com/siteinfo/ted.com>, November 2013.
- [17] Mansoor Alicherry, Muthusrinivasan Muthuprasanna, and Vijay Kumar. High speed pattern matching for network IDS/IPS. In *Network Protocols*,

2006. *ICNP'06. Proceedings of the 2006 14th IEEE International Conference on*, pages 187–196. IEEE, 2006.
- [18] Anchor. MySQL vs. PostgreSQL Comparison.  
[https://www.anchor.com.au/hosting/dedicated/mysql\\_vs\\_postgres](https://www.anchor.com.au/hosting/dedicated/mysql_vs_postgres), 2010.
- [19] Craig A Anderson and Brad J Bushman. Human aggression. *Annual review of psychology*, 53(1):27–51, 2002.
- [20] Eduardo B Andrade, Velitchka Kaltcheva, and Barton Weitz. Self-disclosure on the web: the impact of privacy policy, reward, and company reputation. *Advances in Consumer Research*, 29(1):350–353, 2002.
- [21] Jonathan Anker. Facebook users hate Timeline... and change.  
<http://www.hlntv.com/article/2012/01/31/facebook-timeline-profile-poll-sophos>, May 2012.
- [22] The Internet Archive. Wikipedia. <https://web.archive.org/web/20110606232923/http://www.wikipedia.org/>, June 2011.
- [23] The Internet Archive. Google.  
[https://web.archive.org/web/\\*/http://google.com](https://web.archive.org/web/*/http://google.com), December 2013.
- [24] The Internet Archive. Internet Archive WayBackMachine.  
<https://archive.org/web/>, December 2013.
- [25] The Internet Archive. Youtube.  
[https://web.archive.org/web/\\*/http://youtube.com](https://web.archive.org/web/*/http://youtube.com), December 2013.

- [26] Michael Arrington. Digg To Layoff 37% Of Staff, Product Refocus Imminent. <http://techcrunch.com/2010/10/25/digg-to-lay-off-37-percentof-staff/>, October 2010.
- [27] David A Asch, M Kathryn Jedrzejewski, and Nicholas A Christakis. Response rates to mail surveys published in medical journals. *Journal of clinical epidemiology*, 50(10):1129–1136, 1997.
- [28] Solomon E Asch. Opinions and social pressure. *Readings about the social animal*, pages 17–26, 1955.
- [29] Michael Bächle and Paul Kirchberg. Ruby on Rails. *Software Technology*, December 2007.
- [30] Jenny Bahn. The Insider Source, About. <http://theinsidesource.tumblr.com/about>, 2012.
- [31] Greg Baker. Client-Side Programming. <http://www.cs.sfu.ca/~ggbaker/zju/web/js.html>, 2013.
- [32] James Ball. Verizon court order: Telephone call metadata and what it can show. <http://www.theguardian.com/world/2013/jun/06/phone-call-metadata-information-authorities>, 2013.
- [33] James Ball and Spencer Ackerman. NSA loophole allows warrantless search for US citizens’ emails and phone calls. <http://www.theguardian.com/world/2013/aug/09/nsa-loophole-warrantless-searches-email-calls>, 2013.

- [34] Paul Barrett. Euclidean distance: Raw, normalised, and double-scaled coefficients. Technical report, Cognadev LTD, 2005.
- [35] Adam Barth, Collin Jackson, and John C Mitchell. Robust defenses for cross-site request forgery. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 75–88. ACM, 2008.
- [36] Roy F. Baumeister and Mark R. Leary. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3):497–529, 1995.
- [37] Michael Bernard. Criteria for optimal web design (designing for usability). [http://uwf.edu/ddawson/d3net/documents/web\\_usability/optimal%20web%20design.pdf](http://uwf.edu/ddawson/d3net/documents/web_usability/optimal%20web%20design.pdf), 2003.
- [38] Elisa Bertino and Ravi Sandhu. Database security-concepts, approaches, and challenges. *Dependable and Secure Computing, IEEE Transactions on*, 2(1):2–19, 2005.
- [39] Rob Bertsche and Peter Caruso. Interview with Rob Bertsche and Peter Caruso. Interview, December 2013.
- [40] Eric Bloedorn, Alan D Christiansen, William Hill, Clement Skorupka, Lisa M Talbot, and Jonathan Tivel. Data mining for network intrusion detection: How to get started. Technical report, MITRE Technical Report, 2001.
- [41] Benjamin S Bloom. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6):4–16, 1984.



- [42] Nathalie Bonnardel, Annie Piolat, and Ludovic Le Bigot. The impact of colour on website appeal and users' cognitive processes. *Displays*, 32(2):69–80, 2011.
- [43] Stephen W Boyd and Angelos D Keromytis. SQLrand: Preventing SQL injection attacks. In *Applied Cryptography and Network Security*, pages 292–302. Springer, 2004.
- [44] Jack W Brehm. *Organization Change: A Comprehensive Reader*. Jossey-Bass, 999 Market Street, San Francisco, CA, 94103-1741, 2008.
- [45] Joana Brenner. Pew Internet: Mobile. <http://pewinternet.org/Commentary/2012/February/Pew-Internet-Mobile.aspx>, 2013.
- [46] John Brooke. SUS—a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996.
- [47] Russell Gott Brooker and Todd Michael Schaefer. *Public opinion in the 21st century: let the people speak?* Houghton Mifflin Company, 2006.
- [48] Chad Brooks. What is RAM?  
<http://www.technewsdaily.com/6287-what-is-ram.html>, September 2012.
- [49] Burleson Consulting. Advantages and Disadvantages of Object-Oriented Approach.  
[http://www.dba-oracle.com/t\\_object\\_oriented\\_approach.htm](http://www.dba-oracle.com/t_object_oriented_approach.htm), 2012.
- [50] Kelli S Burns and Richard J Lutz. The function of format: Consumer responses to six on-line advertising formats. *Journal of Advertising*, 35(1):53–63, 2006.

- [51] Daniel Burstein. Marketing research chart: Average website conversion rates, by industry. *MarketingSherpa*, October 2012.
- [52] Margaret Butler. Android: Changing the Mobile Landscape. *Pervasive Computing, IEEE*, 10(1):4–7, 2011.
- [53] LLC Byron Carlson Petri & Kalb. St. Louis charitable 501(c)(3) lawyers. <http://www.bcpklaw.com/Business-Law/Charitable-501-c-3.shtml>, 2013.
- [54] CaldwellYSR. Aptana Studio 3. <http://codersreviews.wordpress.com/2012/06/27/aptana-studio/>, June 2012.
- [55] Franco Callegati, Walter Cerroni, and Marco Ramilli. Man-in-the-Middle Attack to the HTTPS protocol. *Security & Privacy, IEEE*, 7(1):78–81, 2009.
- [56] Flavián Carlos, Gurrea Raquel, and Orús Carlos. The relevance of web design for the website success: A heuristic analysis. *University of Zaragoza*, 2008.
- [57] Sean Carlos. 8 reasons to avoid Flash (or Silverlight) like the plague when designing a website. <http://antezeta.com/news/flash-problems>, March 2007.
- [58] Jamie Carter. Big data and how it will change your world. *MSN Innovation*, pages 1–6, November 2013.
- [59] Eve M Caudill and Patrick E Murphy. Consumer online privacy: Legal and ethical issues. *Journal of Public Policy & Marketing*, pages 7–19, 2000.

- [60] Laure Cetin. The SAP Community Network: How to use gamification to increase engagement. <http://uberdruber.com/index.php/en/social-media/160-the-sap-community-network-how-to-use-gamification-to-increase-engagement>, 2013.
- [61] Dave Chaffey. Display advertising clickthrough rates. *Smart Insights*, 2013.
- [62] Rocky KC Chang. Defending against flooding-based distributed denial-of-service attacks: A tutorial. *Communications Magazine, IEEE*, 40(10):42–51, 2002.
- [63] Shuo Chen, Ziqing Mao, Yi-Min Wang, and Ming Zhang. Pretty-bad-proxy: An overlooked adversary in browsers’ HTTPS deployments. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 347–359. IEEE, 2009.
- [64] Chosun.com. Real-name online registration to be scrapped. [http://english.chosun.com/site/data/html\\_dir/2011/12/30/2011123001526.html](http://english.chosun.com/site/data/html_dir/2011/12/30/2011123001526.html), 2011.
- [65] Elaine Chou. Redesigning a large and complex website: how to begin, and a method for success. In *Proceedings of the 30th annual ACM SIGUCCS conference on User services*, pages 22–28. ACM, 2002.
- [66] Elizabeth G Cohen and Susan S Roper. Modification of interracial interaction disability: An application of status characteristic theory. *American Sociological Review*, pages 643–657, 1972.
- [67] Geoffrey L Cohen, Joshua Aronson, and Claude M Steele. When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, 26(9):1151–1164, 2000.

- [68] Perkins Coie. Security breach notification chart.  
[http://www.perkinscoie.com/files/upload/PS\\_13\\_10\\_SecurityBreachNotificationLawChart.pdf](http://www.perkinscoie.com/files/upload/PS_13_10_SecurityBreachNotificationLawChart.pdf), October 2013.
- [69] Federal Trade Commission. Fair Information Practice Principles.  
<http://www.ftc.gov/reports/privacy3/fairinfo.shtm>, 2012.
- [70] Traverse Global Communications. Non-profit web rewards program features.  
<http://www.nonprofitwebhosting.org/features.html>, 2013.
- [71] CompareVPS. Compare VPS hosting plans.  
[http://www.comparevps.com/?mem\\_min=&disk\\_min=50](http://www.comparevps.com/?mem_min=&disk_min=50), 2013.
- [72] CompareVPS. Selecting a VPS plan.  
<http://www.comparevps.com/guide/select-vps/>, 2013.
- [73] US Congress. Digital Millennium Copyright Act. *Public Law*, (105-304):112, 1998.
- [74] Celina Conner. PROS and CONS of using Twitter Bootstrap. <http://www.thedesignwork.com/pros-and-cons-of-using-twitter-bootstrap/>, February 2013.
- [75] Lucian Constantin. Hackers increasingly target shared web hosting servers for use in mass phishing attacks. *PCWorld*, April 2013.
- [76] Josh Constine. How Big Is Facebook’s Data? 2.5 Billion Pieces of Content and 500+ Terabytes Ingested Every Day. August 2012.
- [77] ConvinceMe. convinceme: start a debate. <http://www.convinceme.net/>, November 2007-2012.

- [78] Copilevitz & Canter, LLC. Effective, experienced counsel for non-profit organizations. <http://www.exempttaxlaw.com/>, 2013.
- [79] Joshua Correll, Steven J. Spencer, and Mark P. Zanna. An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, 40:350–356, 2004.
- [80] Crazy Egg. Crazyegg See Where People Click. <http://www.crazyegg.com>, December 2013.
- [81] Thomas Crook, Brian Frasca, Ron Kohavi, and Roger Longbotham. Seven pitfalls to avoid when running controlled experiments on the web. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1105–1114. ACM, 2009.
- [82] Kevin Crowston, Hala Annabi, and James Howison. Defining open source software project success. In *Proceedings of the International Conference on Information Systems*, ICIS 2003, 2003.
- [83] Yanqing Cui and Virpi Roto. How People Use the Web on Mobile Devices. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 905–914, New York, NY, USA, 2008. ACM.
- [84] Dianne Cyr, Milena Head, and Hector Larios. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies*, 68(1):1–21, 2010.
- [85] Joan Daemen and Vincent Rijmen. *The design of Rijndael: AES-the advanced encryption standard*. Springer, 2002.

- [86] Lincoln Dahlberg. The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4):615–633, 2001.
- [87] Lynn Dailey. Navigational web atmospherics: Explaining the influence of restrictive navigation cues. *Journal of Business Research*, 57(7):795–803, 2004.
- [88] Brittany Darwell. Where are Facebook’s mobile-only users coming from? Probably not iPhones.  
<http://www.insidefacebook.com/2012/08/01/where-are-facebooks-mobile-only-users-coming-from-probably-not-iphones/>, August 2012.
- [89] Debate.org. Debate.org. <http://www.debate.org/>, November 2013.
- [90] DebatePolitics. DebatePolitics. <http://www.debatepolitics.com/>, November 2004-2013.
- [91] Debatewise. debatewise: Where Great Minds Differ.  
<http://debatewise.org/>, November 2013.
- [92] Hans Delfs and Helmut Knebl. *Introduction to Cryptography: Principles and Applications*. Springer, 2002.
- [93] DemandBase. Why B2B is now embracing display advertising: Intent vs. interest. <http://www.demandbase.com/company-targeted-advertising/view/why-b2b-is-now-embracing-display-advertising-intent-vs-interest>, 2013.
- [94] Cathy Derecki. Incremental Process for Eventual Web ReDesign: Start with the Data. <http://cderecki.umwblogs.org/2010/07/15/incremental->

process-for-eventual-web-re-design-start-with-the-data/, July 2010.

- [95] Yvo Desmedt. Man-in-the-middle attack. In *Encyclopedia of Cryptography and Security*, pages 368–368. Springer, 2005.
- [96] Sebastian Deterding. Gamification: Designing for motivation. *interactions*, 19(4):14–17, July 2012.
- [97] Morton Deutsch and Harold B Gerard. A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, 51(3):629, 1955.
- [98] Google Developers. Web metrics: Size and number of resources. <https://developers.google.com/speed/articles/web-metrics>, March 2012.
- [99] Devina. Top 5 best online payment gateways—online payment processors. <http://www.era.la/top-5-best-online-payment-gateways-online-payment-processors.html>, 2012.
- [100] Ravi Dhar. Consumer preference for a no-choice option. *Journal of Consumer Research*, 24(2):215–31, September 1997.
- [101] ESV Digital. Mobile Behaviors Big Demographic Differences to Understand. <http://www.esvdigital.com/blog/2013/09/03/mobile-behaviorsbig-demographic-differences-to-understand/>, September 2013.
- [102] Soussan Djmasbi. Interview with Professor Soussan Djmasbi. Interview, November 2013.

- [103] Domparison. Cheapest .ORG domain registration.  
<http://www.domparison.com/domain-name-price-comparison/index.php?want=register&ext=org>, 2013.
- [104] Karen M. Douglas and Craig McGarty. Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British Journal of Social Psychology*, 40(3):399–416, 2001.
- [105] DreamHost. Non-profit discount.  
<http://wiki.dreamhost.com/Non-profit>, 2013.
- [106] Bobby Duffy, Kate Smith, George Terhanian, and John Bremer. Comparing data from online and face-to-face surveys. *International Journal of Market Research*, 47(6):615, 2005.
- [107] Vu Xuan Duong. *A Proposal of a Cross-Browser User Tracking Method with Browser Fingerprint*. PhD thesis, Keio University, 2011.
- [108] Elizabeth Dwoskin. Twitter’s data business proves lucrative. *The Wall Street Journal*, pages 1–4, October 2013.
- [109] Dylan. Thanks for the gold!  
<http://blog.reddit.com/2013/10/thanks-for-gold.html>, October 2013.
- [110] Julia B Earp, Annie I Antón, Lynda Aiman-Smith, and William H Stufflebeam. Examining internet privacy policies within the context of user privacy values. *Engineering Management, IEEE Transactions on*, 52(2):227–237, 2005.
- [111] Peter Eckersley. How unique is your web browser? In *Privacy Enhancing Technologies*, pages 1–18. Springer, 2010.



- [112] Jim Edwards. Reddit CEO admits ‘We’re still in the red’. *Business Insider*, July 2013.
- [113] Jim Edwards. Yes, your credit card company is selling your purchase data to online advertisers. *Business Insider*, pages 1–1, April 2013.
- [114] Jay Ehret. Know your website hosting options. <http://themarketingspot.com/2011/02/know-your-website-hosting-options.html>.
- [115] Gary Eisenstein. In-house vs. outsourced web hosting. <http://blog.falcon-software.com/2010/08/24/in-house-vs-outsourcing-your-web-hosting/>, August 2010.
- [116] John Elder. Top 10 data mining mistakes. [http://www.sas.com/news/sascom/2010q3/column\\_tech.html](http://www.sas.com/news/sascom/2010q3/column_tech.html), December 2013.
- [117] Electronic Frontier Foundation. Encryption basics. <https://ssd EFF.org/tech/encryption>, December 2013.
- [118] Sanne Elling, Leo Lentz, and Menno de Jong. Users’ abilities to review web site pages. *Journal of Business and Technical Communication*, 26(2):171–201, 2012.
- [119] Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems*. Addison-Wesley, 6 edition, 2011.
- [120] eMarketer. US Smartphone OS Race Still Close, as Men, Younger Users Favor Android. <http://www.emarketer.com/Article/US-Smartphone-OS-Race-Still-Close-Men-Younger-Users-Favor-Android/1009961>, June 2013.

- [121] EnterpriseDB. Why move from Oracle's MySQL to Postgres Plus?  
<http://www.enterprisedb.com/solutions/mysql-vs-postgresql/why-move-mysql-postgres-plus>, 2013.
- [122] William M Epstein. Response bias in opinion polls and American social welfare. *The Social Science Journal*, 43(1):99–110, 2006.
- [123] Charlotte Erdmann. 10 Tips to Improve Mobile Security. [http://www.internetevolution.com/author.asp?section\\_id=2683&doc\\_id=261043](http://www.internetevolution.com/author.asp?section_id=2683&doc_id=261043), March 2013.
- [124] expo09. Advantages and disadvantages of colocation. <http://expo09.hubpages.com/hub/Advantages-and-Disadvantages-of-Colocation>, July 2011.
- [125] Antoine De Saint Exupery. Virtual Dictionary - Famous Quotes. [http://quotes.vdictionary.com/aquotes-Antoine\\_De\\_Saint\\_Exupery.htm](http://quotes.vdictionary.com/aquotes-Antoine_De_Saint_Exupery.htm), December 2013.
- [126] Facebook. How to build your audience on facebook.  
<https://www.facebook.com/advertising/how-it-works>, 2013.
- [127] FactCheck.org. About Us.  
<http://www.factcheck.org/spindetectors/about/>, November 2013.
- [128] Wei Fang. Using google analytics for improving library website content and design: a case study. *Library Philosophy and Practice*, 2007.
- [129] The Digital FAQ. Overselling vs. overloading: What is overselling? (part 1).  
<http://www.digitalfaq.com/guides/webhosting/overselling-servers-hosting.htm>, 2002-2013.

- [130] Farlex. The free dictionary.  
<http://financial-dictionary.thefreedictionary.com/>, 2008.
- [131] Kevin Fayle. Understanding the legal issues for social networking sites and their users. <http://technology.findlaw.com/modern-law-practice/understanding-the-legal-issues-for-social-networking-sites-and.html>, 2008.
- [132] Ad Feelders, H Daniels, and Marcel Holsheimer. Methodological and practical aspects of data mining. *Information & Management*, 37(5):271–281, 2000.
- [133] Steven Fein and Steven J. Spencer. Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73(1):31–44, 1997.
- [134] Philip M Fernbach, Todd Rogers, Craig R Fox, and Steven A Sloman. Political extremism is supported by an illusion of understanding. *Psychological Science*, 24(6):939–946, 2013.
- [135] Roy T. Fielding and Gail Kaiser. The Apache HTTP server project. *Internet Computing, IEEE*, 1(4):88–90, 1997.
- [136] FindLaw. What is a subpoena. <http://litigation.findlaw.com/going-to-court/what-is-a-subpoena.html>, 2013.
- [137] FireBubble. A guide to website hosting (compare different types of web hosting). <http://www.firebubble.co.uk/website-hosting-guide/>, 2013.

- [138] Carlos Flavián, Miguel Guinalíu, and Raquel Gurrea. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1):1–14, 2006.
- [139] Brian Fling. *Mobile Design and Development: Practical Concepts and Techniques for Creating Mobile Sites and Web Apps*. O’Reilly Media, 1 edition, August 2009.
- [140] Network for Good. Online fundraising software and tools for nonprofits. <http://www1.networkforgood.org/digitalgivingindex>, 2013.
- [141] ForandAgainst. For and Against: The Best Place to State your Case. <http://www.forandagainst.com/>, November 2007.
- [142] Ben Foster. How Many Users on Facebook. <http://www.benphoster.com/facebook-user-growth-chart-2004-2010/>, April 2013.
- [143] Electronic Frontier Foundation. Do not track. <https://www EFF.org/issues/do-not-track>, 2013.
- [144] Megan Friedman. Digg Users Lash Out At New Format, Join Forces With Reddit. <http://newsfeed.time.com/2010/08/30/digg-users-lash-out-at-new-format-join-forces-with-reddit/>, August 2010.
- [145] Rich G. 12 advantages of outsource hosting vs. in house hosting. <http://www.inetu.net/about/server-smarts-blog/may-2012/12-advantages-of-outsource-hosting-vs-in-house-hos>, May 2012.
- [146] Jesse Chandler Gabriele Paolacci. Running Experiments on Amazon Mechanical Turk. *Judgement and Decision Making*, 5(5):411–419, July 2010.

- [147] Sean Gallagher. How the NSA's MUSCULAR tapped Google's and Yahoo's private networks.  
<http://arstechnica.com/information-technology/2013/10/how-the-nsas-muscular-tapped-googles-and-yahoos-private-networks/>, 2013.
- [148] Pravin Ganore. Colocation services explained. [https://www.ibm.com/developerworks/community/blogs/df1b71df-7b8b-4063-89ed-620f7a0a5560/entry/colocation\\_services\\_explained2?lang=en](https://www.ibm.com/developerworks/community/blogs/df1b71df-7b8b-4063-89ed-620f7a0a5560/entry/colocation_services_explained2?lang=en), September 2012.
- [149] Wolfgang A Gehrke and Franco Milicchio. *Distributed services with openafs: for enterprise and education*. Springer, 2007.
- [150] Carole A George. Usability testing and design of a library website: An iterative approach. *OCLC Systems & Services*, 21(3):167–180, 2005.
- [151] Mariella Di Giacomo. MySQL: Lessons Learned on a Digital Library.  
<http://www.computer.org/csdl/mags/so/2005/03/s3010.pdf>, June 2005.
- [152] Zoltán Gócza. Myth #11: You need to redesign your website periodically.  
<http://uxmyths.com/post/712537920/myth-you-need-to-redesign-your-website-periodically>, December 2010.
- [153] Google. Features. <http://www.google.com/analytics/features/>, December 2013.
- [154] Google. Google ad grants program details.  
<http://www.google.com/grants/details.html>, 2013.
- [155] Google. How AdWords works. [https://support.google.com/adwords/answer/2497976?hl=en&ref\\_topic=3121763](https://support.google.com/adwords/answer/2497976?hl=en&ref_topic=3121763), 2013.

- [156] Google. Manage costs. <http://www.google.com/ads/displaynetwork/manage-your-ads/manage-costs.html>, 2013.
- [157] Google. Overview of comments and discussions - Drive Help. <https://support.google.com/drive/answer/65129?hl=en>, 2013.
- [158] Kevin Gosztola. FISA court: Congress, phone companies could have challenged NSA's bulk data collection (but didn't). <http://dissenter.firedoglake.com/2013/09/17/fisa-court-would-prefer-americans-be-outrage-at-phone-companies-congress-for-nsa-surveillance-not-them/>, 2013.
- [159] David Gowans. Choosing a webhost. <http://www.freewebmasterhelp.com/articles/webhost>, 1999 - 2003.
- [160] Chris Goward. An Introduction to Evolutionary Site Redesign with WiderFunnel's Blog. <http://www.widerfunnel.com/case-study/an-introduction-to-evolutionary-site-redesign-and-widerfunnels-blog-redesign>, December 2012.
- [161] Mark A Graber, Donna M D Alessandro, and Jill Johnson-West. Reading level of privacy policies on internet health web sites. *Journal of Family Practice*, 51(7):642-642, 2002.
- [162] Ami Grant. Super cookies raise privacy concerns. <http://www.thesearchagents.com/2011/08/super-cookies-raise-privacy-concerns/>, August 2011.

- [163] Grassroots. Free web hosting services for nonprofits and charities. <http://www.grassroots.org/free-website-tools-bluehost-web-hosting>, 2013.
- [164] Robert Gravelle. Review of Aptana Studio 3. <http://www.htmlgoodies.com/beyond/webmaster/toolbox/review-of-aptana-studio-3.html#fbid=1ZMFaXbpYpu>, 2012.
- [165] Hank Green and Alan Lastufka. Community-run music label DFTBA thrives. <http://youtube-global.blogspot.com/2010/12/community-run-music-label-dftba-thrives.html>, 2010.
- [166] Glenn Greenwald and James Ball. The top secret rules that allow NSA to use US data without a warrant. <http://www.theguardian.com/world/2013/jun/20/fisa-court-nsa-without-warrant>, 2013.
- [167] Glenn Greenwald and Ewen MacAskill. NSA prism program taps in to user data of Apple, Google and others. <http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>, 2013.
- [168] Rick Grehan. Pillars of Python: Django Web Framework. <http://www.infoworld.com/d/application-development/pillars-python-django-web-framework-168643>, August 2011.
- [169] Rick Grehan. Pillars of Python: Web.py Web Framework. <http://www.infoworld.com/d/application-development/pillars-python-webpy-web-framework-169072>, August 2011.
- [170] H Paul Grice. Logic and conversation. *Syntax and semantics*, 3:41–58, 1975.

- [171] Mario Guimaraes and Meg Murray. Overview of intrusion detection and intrusion prevention. In *Proceedings of the 5th Annual Conference on Information Security Curriculum Development*, pages 44–46. ACM, 2008.
- [172] WG Halfond, Jeremy Viegas, and Alessandro Orso. A classification of SQL-injection attacks and countermeasures. In *Proceedings of the IEEE International Symposium on Secure Software Engineering, Arlington, VA, USA*, pages 13–15, 2006.
- [173] William GJ Halfond and Alessandro Orso. AMNESIA: Analysis and Monitoring for NEutralizing SQL-Injection Attacks. In *Proceedings of the 20th IEEE/ACM international Conference on Automated software engineering*, pages 174–183. ACM, 2005.
- [174] Marty Hall. *Core Servlets and JavaServer Pages (JSP)*, chapter 9. Prentice Hall PTR/Sun Microsystems Press, 1 edition, 2000.
- [175] Benjamin Halpert. Mobile Device Security. In *Proceedings of the 1st Annual Conference on Information Security Curriculum Development*, InfoSecCD ’04, pages 99–101, New York, NY, USA, 2004. ACM.
- [176] Il-Horn Hann, Jeff Roberts, Sandra Slaughter, and Roy Fielding. Why do developers contribute to open source projects? First evidence of economic incentives. In *2nd Workshop on Open Source Software Engineering, Orlando, FL*, 2002.
- [177] Guy Harrison. NoSQL and Document-Oriented Databases. *Database Trends and Applications*, pages 1–1, November 2010.



- [178] Jeff Hendrickson. Text ads: Unobtrusive advertising. <http://www.jeffhendricksondesign.com/text-ads-unobtrusive-advertising/>, January 2006.
- [179] Mary Henle and Solomon E. Asch. *Documents of Gestalt Psychology*, volume 1. University of California Press, 1961.
- [180] Marc J Hetherington. Resurgent mass partisanship: The role of elite polarization. *American Political Science Review*, 95(3):619–632, 2001.
- [181] Frederick J. Hirsch. Introducing SSL and certificates using SSLeay. <http://home.comcast.net/~fjhirsch/Papers/wwwj/article.html>, 1997.
- [182] Chia-Huan Ho and Karen Swan. Evaluating online conversation in an asynchronous learning environment: An application of grice’s cooperative principle. *The Internet and Higher Education*, 10(1):3–14, 2007.
- [183] Matthew Hornsey, Louise Majkut, Deborah Terry, and Blake McKimmie. On being loud and proud: Non-conformity and counter-conformity to group norms. *British Journal of Social Psychology*, 42(3):319–335, 2003.
- [184] Peter Hornyack, Seungyeop Han, Jaeyeon Jung, Stuart Schechter, and David Wetherall. ”these Aren’t the Droids You’re Looking For”: Retrofitting Android to Protect Data from Imperious Applications. <http://research.microsoft.com/pubs/149596/AppFence.pdf>, 2011.
- [185] Privacy HTTP Cookies: Standards and Politics. David M. Kristol. *ACM Transactions on Internet Technology*, 1(2), November 2001.
- [186] Abine, Inc. How you can be tracked online. <http://abine.com/tracking.php>, 2012.

- [187] Philip G Inglesant and M Angela Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 383–392. ACM, 2010.
- [188] Open Source Initiative. The open source definition.  
<http://opensource.org/osd>, 2013.
- [189] Applied Innovations. A quick comparison of in-house versus hosted web solutions. <http://www.appliedi.net/blog/a-quick-comparison-of-in-house-versus-hosted-web-solutions/>, January 2013.
- [190] International Debate Education Association. Idea Beta.  
<http://idebate.org/>, November 2013.
- [191] InterWorking Labs. Man in the middle attacks.  
<http://iwl.com/solutions/man-middle-attacks>, December 2013.
- [192] Luke P. Issac. Nginx Vs Apache: Nginx Basic Architecture and Scalability.  
<http://www.thegeekstuff.com/2013/11/nginx-vs-apache/>, November 2013.
- [193] Solid IT. DB-Engines Ranking of Document Stores.  
<http://db-engines.com/en/ranking/document+store>, December 2013.
- [194] Solid IT. DB-Engines Ranking of Object Oriented DBMS.  
<http://db-engines.com/en/ranking/object+oriented+dbms>, December 2013.
- [195] Solid IT. DB-Engines Ranking of Relational DBMS.  
<http://db-engines.com/en/ranking/relational+dbms>, December 2013.

- [196] Solid IT. Popularity changes per category, December 2013.  
[http://db-engines.com/en/ranking\\_categories](http://db-engines.com/en/ranking_categories), December 2013.
- [197] Sheena S Iyengar and Mark R Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995, 2000.
- [198] Bryan Morgan James Goodwill. *Developing Java Servlets*, chapter 5. Sams Publishing, 2 edition, May 2001.
- [199] Daniel Jarrett. Choosing a host.  
[http://www.4webhelp.net/tutorials/misc/choose\\_host.php](http://www.4webhelp.net/tutorials/misc/choose_host.php), May 2012.
- [200] Sirrka L. Jarvenpaa and Peter A. Todd. Consumer reactions to electronic shopping on the world wide web. *International Journal of Electronic Commerce*, 1(2):59–88, 1996.
- [201] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: An evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 471–478. ACM, 2004.
- [202] Joe Jensen. What type of web hosting is best for you?  
<http://networkbits.net/web/web-hosting-types/>, September 2008.
- [203] Mark Johanson. Facebook Changes: Users Overwhelmingly 'Dislike' new News Feed and Ticker. <http://www.ibtimes.com/facebook-changes-users-overwhelmingly-dislike-new-news-feed-ticker-317124>, September 2011.

- [204] Michael Johnston. How to monetize a website: The 5 minute monetization guide. <http://monetizepros.com/blog/2013/which-monetization-methods-will-make-the-most-money-for-your-website/>, 2013.
- [205] Paul Johnston. Authentication and session management on the web. *GIAC Security Essentials Certification Practical Assignment Version 1.4b*, November 2004.
- [206] Karine Jolly. Web Redesign on a Dime. <http://www.universitybusiness.com/article/web-redesign-dime>, April 2009.
- [207] Darren Jones. Rails or Sinatra: The Best of Both Worlds? <http://www.sitepoint.com/rails-or-sinatra-the-best-of-both-worlds/>, January 2012.
- [208] Zach Jones. Web storage security. *Technical Insight, True Stories of the TRC, Vulnerabilities*, May 2013.
- [209] Konstantin Käfer. Cross site request forgery. *Hasso-Plattner-Institut*, December 2008.
- [210] Martin F. Kaplan and Charles E. Miller. Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. *Journal of Personality and Social Psychology*, 53(2):306–313, 1987.
- [211] Kate Kaye. Mastercard, AmEx Quietly Feed Data to Advertisers. *Ad Age*, pages 1–3, April 2013.

- [212] George L Kelling and Catherine M Coles. *Fixing broken windows: Restoring order and reducing crime in our communities*. Free Press, 1998.
- [213] Chris “KeyserSosa”. This was a triumph.  
<http://blog.reddit.com/2010/07/it-was-triumph.html>, July 2010.
- [214] Muneo Kitajima, Noriyuki Kariya, Hideaki Takagi, and Yongbing Zhang. Evaluation of website usability using Markov chains and latent semantic analysis. *IEICE transactions on communications*, (4):1467–1475, 2005.
- [215] Dave Klein. How to decide: Mobile websites vs. mobile apps. *Inspire*, February 2012.
- [216] Shawn Knight. Get paid to sell your personal data to companies with Handshake. *TechSpot*, pages 1–1, September 2013.
- [217] Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M Henne. Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1):140–181, 2009.
- [218] Sander L. Koole, Karianne Smeets, Ad Van Knippenberg, and Ap Dijksterhuis. The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, 77(1):111–125, 1999.
- [219] Denis Krienbhl. Flask vs. Django vs. Pyramid vs. Plone.  
<http://blog.stacktrace.ch/post/49178654214>, April 2013.
- [220] Jennifer Kyrnin. Don’t Use Flash This Way. [http://webdesign.about.com/od/flash/qt/dont\\_use\\_flash\\_this\\_way.htm](http://webdesign.about.com/od/flash/qt/dont_use_flash_this_way.htm), November 2013.

- [221] Karim R Lakhani and Eric Von Hippel. How open source software works: "free" user-to-user assistance. *Research Policy*, 32(6):923–943, 2003.
- [222] Cliff Lampe, Nicole B Ellison, and Charles Steinfield. Changes in use and perception of Facebook. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 721–730. ACM, 2008.
- [223] Noam Lapidot-Lefler and Azy Barak. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2):434–443, 2012.
- [224] Erik Larkin. Browser fingerprints: A big privacy threat. *PCWorld*, March 2010.
- [225] Felix Lau, Stuart H Rubin, Michael H Smith, and Ljiljana Trajkovic. Distributed denial of service attacks. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 3, pages 2275–2280. IEEE, 2000.
- [226] Paul J. Lavrakas and Ana Villar. *Encyclopedia of Survey Research Methods*. Sage Publications, Inc, 1 edition, 2008.
- [227] George Lawton. Is It Finally Time to Worry about Mobile Malware? *Computer*, 41(5):12–14, 2008.
- [228] Geoffrey C Layman, Thomas M Carsey, and Juliana Menasce Horowitz. Party polarization in American politics: Characteristics, causes, and consequences. *Annual Review of Political Science*, 9:83–110, 2006.

- [229] Martin Lea, Russell Spears, and Daphne de Groot. Knowing me, knowing you: Anonymity effects on social identity processes within groups. *Personality and Social Psychology Bulletin*, 27(5):526–537, 2001.
- [230] Neal Leavitt. Mobile Phones: The Next Frontier for Hackers? *Computer*, 38(4):20–23, 2005.
- [231] Nevada Legislature. Nevada Revised Statutes 603a: Security of personal information. <http://www.leg.state.nv.us/nrs/nrs-603a.html>, 2011.
- [232] Timo Lehtonen, Said Benamar, Vesa Laamanen, Ilkka Luoma, Olli Ruotsalainen, Jaakko Salonen, and Tommi MikkonenP. Towards User-Friendly Mobile Browsing. In *Proceedings of the 2Nd International Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications*, AAA-IDEA '06, New York, NY, USA, 2006. ACM.
- [233] Sointu Leikas, Jan-Erik Lönnqvist, and Markku Verkasalo. Persons, situations, and behaviors: Consistency and variability of different behaviors in four interpersonal situations. *Journal of Personality and Social Psychology*, 2012.
- [234] Rasmus Lerdorf. The no-framework PHP MVC framework. <http://toys.lerdorf.com/archives/38-The-no-framework-PHP-MVC-framework.html>, February 2006.
- [235] Josh Lerner and Jean Tirole. Some simple economics of open source. *The Journal of Industrial Economics*, 50(2):197–234, 2002.
- [236] Reuven M. Lerner. At the Forge: Twitter Bootstrap. *Linux J.*, 2012(218), June 2012.

- [237] James R Lewis. Sample sizes for usability studies: Additional considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2):368–378, 1994.
- [238] Emman Lijesta. User experience: Unobtrusive website ads. <http://www.clickbankphilippines.com/2010/10/user-experience-unobtrusive-website-ads.html>, October 2010.
- [239] LimeService. Data Protection Statement. <http://www.limeservice.com/en/component/content/article/21-english/general-content/39-data-protection-statement>, February 2011.
- [240] Michael W Link, Michael P Battaglia, Martin R Frankel, Larry Osborn, and Ali H Mokdad. Reaching the US cell phone generation: Comparison of cell phone survey results with an ongoing landline telephone survey. *Public Opinion Quarterly*, 71(5):814–839, 2007.
- [241] Patricia W Linville. The complexity–extremity effect and age-based stereotyping. *Journal of Personality and Social Psychology*, 42(2):193, 1982.
- [242] Alex X Liu, Jason M Kovacs, C-T Huang, and Mohamed G Gouda. A secure cookie protocol. In *Computer Communications and Networks, 2005. ICCCN 2005. Proceedings. 14th International Conference on*, pages 333–338. IEEE, 2005.
- [243] Chang Liu and Kirk P. Arnett. Exploring the factors associated with web site success in the context of electronic commerce. *Information & Management*, 38(1):23–33, 2000.



- [244] McKenna Long & Aldridge LLP. California Online Privacy Protection Act of 2003. [http://www.mckennalong.com/media/site\\_files/194\\_California%20Online%20Privacy%20Protection%20Act%20of%202003.pdf](http://www.mckennalong.com/media/site_files/194_California%20Online%20Privacy%20Protection%20Act%20of%202003.pdf), 2004.
- [245] Gerald L. Lohse and Peter Spiller. Internet retail store design: How the user interface influences traffic and sales. *Journal of Computer-Mediated Communication*, 5(2):0–0, 1999.
- [246] Bit Zesty Ltd. Document Stores vs Relational Databases for Domain Modelling. <http://bitzesty.com/blog/2010/01/document-stores-vs-relational-databases-for-domain-modelling/>, January 2010.
- [247] Shanshan Ma. 10 Ways Mobile Sites Are Different from Desktop Web Sites. <http://www.uxmatters.com/mt/archives/2011/03/10-ways-mobile-sites-are-different-from-desktop-web-sites.php>, March 2011.
- [248] Ewen MacCaskill and Gabriel Dance. NSA files: Decoded. <http://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded>, 2013.
- [249] David P MacKinnon and Angela Lapin. Effects of alcohol warnings and advertisements: A test of the boomerang hypothesis. *Psychology and Marketing*, 15(7):707–726, 1998.
- [250] Steve MacLaughlin. How much money is raised through online giving? *npENGAGE*, December 2011.
- [251] Steve MacLaughlin. Charitable giving report how nonprofit fundraising performed in 2012. Technical report, Blackbaud, February 2013.

- [252] Alan Mandel. Civil Good. [http://civilgood.org/info/index\\_files/Civil%20Good%20Detail%20Document-29.pdf](http://civilgood.org/info/index_files/Civil%20Good%20Detail%20Document-29.pdf), July 2012.
- [253] MapBox. Mapbox — design & publish beautiful maps. <https://www.mapbox.com/>, 2013.
- [254] Ethan Marcotte. Responsive Web Design. <http://alistapart.com/article/responsive-web-design>, May 2010.
- [255] MarketingCharts. The Demographics of “Cell-Mostly” Internet Users. <http://www.marketingcharts.com/wp/online/the-demographics-of-cell-mostly-internet-users-36730/>, September 2013.
- [256] Patrick M Markey and Shannon M Wells. Interpersonal Perception in Internet Chat Rooms. *Journal of Research in Personality*, 36(2):134–146, 2002.
- [257] Ian Mason. Web hosting types: Advantages and disadvantages. <http://www.tcconcepts.com/web-hosting-options.htm>, 1997-2012.
- [258] Jörg Matthes, Kimberly Rios Morrison, and Christian Schemer. A spiral of silence for some: Attitude certainty and the expression of political minority opinion. *Communication Research*, 37(6):774–800, 2010.
- [259] Erika McCallister. *Guide to protecting the confidentiality of personally identifiable information*. DIANE Publishing, 2010.
- [260] Caroline McCarthy. Angry Digg users flood homepage with Reddit links. [http://news.cnet.com/8301-13577\\_3-20015042-36.html?part=rss&subj=news&tag=2547-1\\_3-0-20](http://news.cnet.com/8301-13577_3-20015042-36.html?part=rss&subj=news&tag=2547-1_3-0-20), August 2010.

- [261] Scott McCoy, Andrea Everard, Dennis Galletta, and Peter Polak. A study of the effects of online advertising: A focus on pop-up and in-line ads. In *Proc. 3rd Ann. Workshop on HCI Research in MIS, Washington DC*. Citeseer, 2004.
- [262] John McDermott. Foursquare selling its location data through ad targeting firm Turn. *Ad Age*, pages 1–4, July 2013.
- [263] Aleecia M Mcdonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. A comparative study of online privacy policies and formats. In *Privacy enhancing technologies*, pages 37–55. Springer, 2009.
- [264] Gary McGraw. Software security. *Security & Privacy, IEEE*, 2(2):80–83, 2004.
- [265] Abhilasha Mehta. Advertising attitudes and advertising effectiveness. *Journal of Advertising Research*, 40(3):67–71, 2000.
- [266] Raj Mehta and Eugene Sivadas. Direct marketing on the internet: An empirical assessment of consumer attitudes. *Journal of Direct Marketing*, 9(3):21 – 32, 1995.
- [267] Alfred J Menezes, Paul C Van Oorschot, and Scott A Vanstone. *Handbook of Applied Cryptography*. CRC, 2010.
- [268] Merriam-Webster. Accuracy.  
<http://www.merriam-webster.com/dictionary/accuracy>, 2013.
- [269] Merriam-Webster. Clarity.  
<http://www.merriam-webster.com/dictionary/clarity>, 2013.

- [270] Theodoric Meyer and Peter Maass. No warrant, no problem: How the government can get your digital data. <http://www.propublica.org/special/no-warrant-no-problem-how-the-government-can-still-get-your-digital-data>, 2013.
- [271] Samuel D. Gosling Michael Buhrmester, Tracy Kwang. Amazon’s Mechanical Turk: A New Source of Inexpensive Yet High-Quality Data? *Perspectives on Psychological Science*, 6(1):3–5, January 2011.
- [272] Graham Clark Michael Rash, Angela Orebaugh. *Intrusion Prevention and Active Response: Deploying Network and Host IPS*. Syngress Publishing, Inc., 2005.
- [273] Microsoft. Man-in-the-middle attack. [http://technet.microsoft.com/en-us/library/gg195821\(v=ocs.14\).aspx](http://technet.microsoft.com/en-us/library/gg195821(v=ocs.14).aspx), June 2011.
- [274] Microsoft. What is a Transaction? [http://msdn.microsoft.com/en-us/library/aa366402\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/aa366402(VS.85).aspx), November 2013.
- [275] Kasia Mikoluk. MySQL vs PostgreSQL: Why MySQL Is Superior To PostgreSQL. <https://www.udemy.com/blog/mysql-vs-postgresql/>, August 2013.
- [276] Elinor Mills. Device identification in online banking is privacy threat, expert says. *CNET*, April 2009.
- [277] George R. Milne and Mary J. Culnan. Strategies for reducing online privacy risks: Why consumers read (or don’t read) online privacy notices. *Journal of Interactive Marketing*, 18(3):15 – 29, 2004.

- [278] Stefan Mischook. What is Flash, when and why to use it?  
[http://www.killersites.com/articles/articles\\_FlashUse.htm](http://www.killersites.com/articles/articles_FlashUse.htm), 2005.
- [279] Jeremy M. Mittman. California set to enact new “do not track” disclosure requirement law for websites. <http://www.jdsupra.com/legalnews/california-set-to-enact-new-do-not-trac-11610/>, 2013.
- [280] Audris Mockus, Roy T Fielding, and James Herbsleb. A case study of open source software development: the Apache server. In *Software Engineering, 2000. Proceedings of the 2000 International Conference on*, pages 263–272. IEEE, 2000.
- [281] Darren B. Moore. *Setting Up Your 501(c)(3) Non-Profit Organization*, chapter 12. Bourland, Wall & Wenzel, P.C., July 2004.
- [282] Joe Mullin. New Numbers From Pandora Show Big Growth But No Profits. *PaidContent*, May 2011.
- [283] Alex Munroe. PHP: a fractal of bad design. <http://me.veekun.com/blog/2012/04/09/php-a-fractal-of-bad-design/>, April 2012.
- [284] Jesse Mutzebaugh. Incremental Website Redesign Series.  
<http://clikfocus.com/blog/incremental-website-redesign-series>, November 2012.
- [285] Lucy Napper, Peter R. Harris, and Tracy Epton. Developing and testing a self-affirmation manipulation. *Self and Identity*, 8(1):45–62, 2009.
- [286] Shannon C. Nelson. *Encyclopedia of Survey Research Methods*. Number ISBN 9781412963947. Sagepub, 2008.

- [287] Trent Nelson and May Chaffin. Common cybersecurity vulnerabilities in industrial control systems. *Control Systems Security Program*, 2011.
- [288] nibusinessinfo.co.uk. Website hosting options.  
<http://www.nibusinessinfo.co.uk/content/house-hosting>, 2012.
- [289] Saverio Niccolini, RG Garroppo, S Giordano, G Risi, and S Ventura. SIP intrusion detection and prevention: recommendations and prototype implementation. In *VoIP Management and Security, 2006. 1st IEEE Workshop on*, pages 47–52. IEEE, 2006.
- [290] Steve Nichols. Big brother is watching. an update on web bugs. *GSEC version 1.2d*, July 2001.
- [291] Jakob Nielsen. Estimating the number of subjects needed for a thinking aloud test. *International journal of human-computer studies*, 41(3):385–397, 1994.
- [292] Richard E Nisbett and Timothy D Wilson. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35(4):250, 1977.
- [293] Elisabeth Noelle-Neumann. Turbulences in the climate of opinion: Methodological applications of the spiral of silence theory. *Public Opinion Quarterly*, 41(2):143–158, 1977.
- [294] Dare Obasanjo. Why Aren’t You Using An Object Oriented Database Management System?  
<http://www.25hoursaday.com/WhyArentYouUsingAnOODBMS.html>, 2001.

- [295] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *ICWSM*, 11:122–129, 2010.
- [296] Jolie O'Dell. eBay launches trend-spotting site based on user data. *ReadWrite*, November 2009.
- [297] Philippe Oechslin. Making a faster cryptanalytic time-memory trade-off. In *Advances in Cryptology-CRYPTO 2003*, pages 617–630. Springer, 2003.
- [298] Massachusetts Office of Consumer Affairs and Business Regulation. 201 CMR 17.00: Standards for the protection of personal information of residents of the commonwealth. <http://www.mass.gov/ocabr/docs/idtheft/201cmr1700reg.pdf>, 2010.
- [299] Bureau of Consumer Protection. Complying with COPPA: Frequently asked questions. <http://www.business.ftc.gov/documents/Complying-with-COPPA-Frequently-Asked-Questions>, 2013.
- [300] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.
- [301] Online Debate Network. ODN The Forge of Reason. <http://www.onlinedebate.net>, November 2003-2012.
- [302] K Palanivel and R Siavkumar. Fuzzy multi-criteria decision-making approach for collaborative recommender systems. *International Journal of Computer Theory and Engineering*, 2(1):57–63, 2010.
- [303] Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. Lurking? Cyclopaths?: A quantitative lifecycle analysis of user behavior in a

- geowiki. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1917–1926. ACM, 2010.
- [304] Zizi Papacharissi. Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2):259–283, 2004.
- [305] Joon S Park and Ravi Sandhu. Secure cookies on the web. *Internet Computing, IEEE*, 4(4):36–44, 2000.
- [306] Clayton Parker. 4 Python Web Frameworks Compared. <http://www.sixfeetup.com/blog/4-python-web-frameworks-compared>, March 2013.
- [307] Charles Pavitt. *Small Group Communication: A Theoretical Approach*. 3 edition, 2002.
- [308] Cem Paya. Privacy and HTTP referer header (2/2). <http://randomoracle.wordpress.com/2013/11/25/privacy-and-http-referer-header-22/>, November 2013.
- [309] Don Peppers and Martha Rogers. Email marketing maximized. *Peppers and Rodgers Group, Stamford, CA*, 2000.
- [310] Nicole Perlroth and Nick Bilton. Mobile Apps Take Data Without Permission. [http://bits.blogs.nytimes.com/2012/02/15/google-and-mobile-apps-take-data-books-without-permission/?\\_r=0](http://bits.blogs.nytimes.com/2012/02/15/google-and-mobile-apps-take-data-books-without-permission/?_r=0).
- [311] Adrian Perrig. Shortcomings of password-based authentication. [https://www.usenix.org/legacy/events/sec2000/full\\_papers/dhamija/dhamija\\_html/node2.html](https://www.usenix.org/legacy/events/sec2000/full_papers/dhamija/dhamija_html/node2.html), September 2000.



- [312] Peter Baldwin, David Price. DebateGraph. <http://debategraph.org/>, November 2013.
- [313] Peter Ditto, Jesse Graham, Jonathan Haidt, Ravi Iyer, Spassena Koleva, Matt Motyl, Gary Sherman, Sean Wojcik. YourMorals.org. <http://yourmorals.org>, December 2013.
- [314] Philippe Beaudette, et al. Fundraising 2010/banner testing. *Wikimedia Meta-Wiki*, 2010.
- [315] Andrew Phippen, L Sheppard, and Steven Furnell. A practical evaluation of web analytics. *Internet Research*, 14(4):284–293, 2004.
- [316] Luke Plant. Why escape-on-input is a bad idea. <http://lukeplant.me.uk/blog/posts/why-escape-on-input-is-a-bad-idea/>, August 2012.
- [317] Julia Plekhanova. Evaluating Web Development Frameworks: Django, Ruby on Rails and CakePHP. Technical report, IBIT, Fox School of Business, September 2009.
- [318] Barbara Poblete and Ricardo Baeza-Yates. A content and structure website mining model. In *Proceedings of the 15th international conference on World Wide Web*, pages 957–958. ACM, 2006.
- [319] Tutorials Point. Web - hosting concepts. [http://www.tutorialspoint.com/web\\_developers\\_guide/web\\_hosting\\_concepts.htm](http://www.tutorialspoint.com/web_developers_guide/web_hosting_concepts.htm), 2013.
- [320] André Pomp. Tracking web users-recent advances. <http://andre-pomp.square7.ch/downloads/Tracking%20Web%20Users%20-%20Recent%20Advances.pdf>, 2012.

- [321] Keith T. Poole and Howard Rosenthal. D-NOMINATE after 10 years: A comparative update to congress: A political-economic history of roll-call voting. *Legislative Studies Quarterly*, 26(1):5–29, 2001.
- [322] David Powers. Which server-side technology should I choose?  
[http://www.adobe.com/devnet/dreamweaver/articles/which\\_serverside\\_technology.html](http://www.adobe.com/devnet/dreamweaver/articles/which_serverside_technology.html), April 2010.
- [323] Jenny Preece. Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology*, 20(5):347–356, 2001.
- [324] Associated Press. Number of active users at facebook over the years.  
<http://finance.yahoo.com/news/number-active-users-facebook-over-years-214600186--finance.html>, October 2012.
- [325] Associated Press. Google asks FISA court to lift gag order on NSA surveillance program. <http://www.cbsnews.com/news/google-asks-fisa-court-to-lift-gag-order-on-nsa-surveillance-program/>, 2013.
- [326] ProCon.org. Pros and Cons of Controversial Issues.  
<http://www.procon.org/>, November 2013.
- [327] Digital Media Law Project. Defamation.  
<http://www.dmlp.org/legal-guide/defamation>, 2008.
- [328] Digital Media Law Project. Fair use.  
<http://www.dmlp.org/legal-guide/fair-use>, 2008.

- [329] Digital Media Law Project. Opinion and fair comment privileges. <http://www.dmlp.org/legal-guide/opinion-and-fair-comment-privileges>, 2008.
- [330] Digital Media Law Project. Proving fault: Actual malice and negligence. <http://www.dmlp.org/legal-guide/proving-fault-actual-malice-and-negligence>, 2008.
- [331] Digital Media Law Project. Copyright claims based on user content. <http://www.dmlp.org/legal-guide/copyright-claims-based-user-content>, 2009.
- [332] Digital Media Law Project. Trademark: User-generated content. <http://www.dmlp.org/legal-guide/trademark-user-generated-content>, 2009.
- [333] Digital Media Law Project. Trademark. <http://www.dmlp.org/legal-guide/trademark>, 2010.
- [334] Digital Media Law Project. Copyright. <http://www.dmlp.org/legal-guide/copyright>, 2011.
- [335] Digital Media Law Project. Immunity for online publishers under the Communications Decency Act. <http://www.dmlp.org/legal-guide/immunity-online-publishers-under-communications-decency-act>, 2011.
- [336] Digital Media Law Project. Risks associated with publication. <http://www.dmlp.org/legal-guide/risks-associated-publication>, 2011.

- [337] Digital Media Law Project. Section 230 of the Communications Decency Act. <http://www.dmlp.org/section-230>, 2011.
- [338] Digital Media Law Project. Protecting yourself against copyright claims based on user content. <http://www.dmlp.org/legal-guide/protecting-yourself-against-copyright-claims-based-user-content>, 2012.
- [339] Joe Purcell. CakePHP vs. CodeIgniter. <http://www.devwebpro.com/cakephp-vs-codeigniter/>, June 2011.
- [340] Bob Kaehms Ric McGredy Rasmus Lerdorf, Kevin Tatroe. Maintaining state. *Programming PHP*, March 2002.
- [341] Eric S Raymond. *The Cathedral & the Bazaar: Musings on Linux and open source by an accidental revolutionary*. O'Reilly, 2008.
- [342] reddit. reddit.com: gold. <http://www.reddit.com/gold/about>, 2013.
- [343] reddit admins. reddit myth busters. [http://blog.reddit.com/2013/08/reddit-myth-busters\\_6.html](http://blog.reddit.com/2013/08/reddit-myth-busters_6.html), August 2013.
- [344] reddit inc. about reddit. <http://www.reddit.com/about/>, December 2013.
- [345] reddit inc. changemyview. <http://www.reddit.com/r/changemyview>, November 2013.
- [346] reddit inc. redditdev. <http://www.reddit.com/r/redditdev/>, December 2013.
- [347] reddit inc. We are sysadmins @ reddit. Ask us anything! [http://www.reddit.com/r/sysadmin/comments/r6zfv/we\\_are\\_sysadmins\\_reddit\\_ask\\_us\\_anything/](http://www.reddit.com/r/sysadmin/comments/r6zfv/we_are_sysadmins_reddit_ask_us_anything/), December 2013.

- [348] Mark B. Reed and Lisa G. Aspinwall. Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion*, 22(2):99–132, 1998.
- [349] Will Reese. Nginx: The High-Performance Web Server and Reverse Proxy. *Linux J.*, 2008(173), September 2008.
- [350] Ruth Rettie. Email marketing: success factors.  
<http://ausweb.scu.edu.au/aw02/papers/refereed/rettie/paper.html>, 2002.
- [351] Kent Roberts. Cloud vs. VPS hosting.  
<http://www.atlantic.net/blog/2013/10/23/cloud-vs-vps-hosting/>, 2013.
- [352] Shelly Rodgers and Esther Thorson. The interactive advertising model: How users perceive and process online ads. *Journal of Interactive Advertising*, 1(1):42–61, 2000.
- [353] Ronald Taylor, Jeffrey S. Tenenbaum & Armand Zottola. Preparing an online social media policy: The top ten legal considerations for your nonprofit. <http://www.jdsupra.com/legalnews/preparing-an-online-social-media-policy-84628/>, 2013.
- [354] Louis Rosenfeld. Stop Redesigning And Start Tuning Your Site Instead.  
<http://uxdesign.smashingmagazine.com/2012/05/16/stop-redesigning-start-tuning-your-site/>, May 2012.
- [355] Jessica Roy. Reddit adopts new ad system, Adzerk, allowing users to up and downvote sidebar ads. *BetaBeat*, March 2013.

- [356] George Rubanenko. Online gift donor profile. Technical report, Blackbaud, May 2006.
- [357] Amir Said. Measuring the strength of partial encryption schemes. In *Image Processing, 2005. ICIIP 2005. IEEE International Conference on*, volume 2, pages II–1126. IEEE, 2005.
- [358] Karen Scarfone and Murugiah Souppaya. Guide to enterprise password management (draft). *NIST Special Publication*, 800:118, 2009.
- [359] Dietram A Scheufele, James Shanahan, and Sei-Hill Kim. Who cares about local politics? Media influences on local political involvement, issue awareness, and attitude strength. *Journalism & Mass Communication Quarterly*, 79(2):427–444, 2002.
- [360] Michael Schmidt. HTML5 web security. *Compass Security Ethical Hacking & Penetration Testing*, December 2011.
- [361] Niklas Schmücker. Web tracking.  
[http://www.snet.tu-berlin.de/fileadmin/fg220/courses/SS11/snet-project/web-tracking\\_schmuecker.pdf](http://www.snet.tu-berlin.de/fileadmin/fg220/courses/SS11/snet-project/web-tracking_schmuecker.pdf), 2011.
- [362] Anne Schneider and Helen Ingram. Social construction of target populations: Implications for politics and policy. *American Political Science Review*, pages 334–347, 1993.
- [363] ScienceDaily. App to Protect Private Data On iOS Devices Finds Almost Half of Other Apps Access Private Data. *ScienceDaily*, June 2013.
- [364] ScienceDaily. Security Comparison: Android Versus Apple. *ScienceDaily*, October 2013.

- [365] Internal Revenue Service. Application for recognition of exemption under section 501(c)(3) of the internal revenue code. Internal Revenue Service Publication Form 1023, 2013.
- [366] Internal Revenue Service. Applying for 501(c)(3) tax-exempt status. Internal Revenue Service Publication 4220, 2013.
- [367] Internal Revenue Service. Tax-exempt status for your organization. Internal Revenue Service Publication 557, 2013.
- [368] Eldar Shafir, Itamar Simonson, and Amos Tversky. Reason-based choice. *Cognition*, 49(1-2):11–36, 1993.
- [369] David K Sherman and Geoffrey L Cohen. Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, 11(4):119–123, 2002.
- [370] David K Sherman and Geoffrey L Cohen. The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, 38:183–242, 2006.
- [371] David K Sherman, Geoffrey L Cohen, Leif D Nelson, A David Nussbaum, Debra P Bunyan, and Julio Garcia. Affirmed yet unaware: exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology*, 97(5):745, 2009.
- [372] Nate Shivar. How to advertise on reddit: My experience & 9 ideas for advertising on reddit. *ShivarWeb*, November 2013.

- [373] SIGCHI Conference on Human Factors in Computing Systems. *Are your participants gaming the system?: screening mechanical turk workers*, Atlanta, Georgia, April 2010. Association for Computing Machinery.
- [374] Silvia. Mobile Web Design. <http://www.onbible.com/info/mobile-web-design/>, October 2013.
- [375] Abe Singer, Warren Anderson, and Rik Farrow. Rethinking password policies. *Usenix*, 38, 2013.
- [376] Eleanor Singer. The Use of Incentives to Reduce Nonresponse in Household Surveys. *Survey nonresponse*, (051):163–177, 2002.
- [377] Alexander Slesarev. Web Application Framework Comparison By Memory Consumption. <http://nuald.blogspot.com/2011/08/web-application-framework-comparison-by.html>, August 2011.
- [378] Jonathan Snook. CodeIgniter vs. CakePHP. [http://snook.ca/archives/php/codeigniter\\_vs\\_cakephp](http://snook.ca/archives/php/codeigniter_vs_cakephp), March 2007.
- [379] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. Flash cookies and privacy. In *AAAI Spring Symposium: Intelligent Information Privacy Management*, 2010.
- [380] Component Based Solutions. Databases: relational vs object vs graph vs document. [http://www.cbsolution.net/techniques/ontarget/databases\\_relational\\_vs\\_object\\_vs](http://www.cbsolution.net/techniques/ontarget/databases_relational_vs_object_vs), May 2011.
- [381] E Isaac Sparling and Shilad Sen. Rating: how difficult is it? In *Proceedings of the fifth ACM conference on Recommender systems*, pages 149–156. ACM, 2011.



- [382] Kevin Spett. Cross-site scripting. *SPI Labs*, pages 1–20, 2005.
- [383] Alfred Spiessens. Patterns of safe collaboration.  
[http://www.evoluware.eu/fsp\\_thesis.pdf](http://www.evoluware.eu/fsp_thesis.pdf), 2007.
- [384] Lee Sproull and Sara Kiesler. Reducing social context cues: Electronic mail in organizational communication. *Management Science*, 32(11):1492–1512, 1986.
- [385] Penn State. Lesson 5: Types of Database Management Systems.  
<http://archive.is/VVkhs>, 2008.
- [386] Gina Marie Stevens. *Data Security Breach Notification Laws*. Congressional Research Service, 2012.
- [387] Stefan Stieger, Ulf-Dietrich Reips, and Martin Voracek. Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology*, 58(11):1653–1660, 2007.
- [388] David Matthias Stöckl. Programming Languages for the Web 2011.  
[http://www.blackbam.at/blackbams-blog/wp-content/uploads/2011/09/programming\\_languages\\_for\\_the\\_web\\_2011.pdf](http://www.blackbam.at/blackbams-blog/wp-content/uploads/2011/09/programming_languages_for_the_web_2011.pdf), 2011.
- [389] Fritz Strack, Norbert Schwarz, Herbert Bless, Almut Kübler, and Michaela Wänke. Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology*, 23(1):53–62, 1993.
- [390] John Suler. *The Psychology of Cyberspace*. True Center Publishing, 16 North Franklin Street Doylestown, Pennsylvania, 18901, 2001-2004.

- [391] Wayne C Summers and Edward Bosworth. Password policy: the good, the bad, and the ugly. In *Proceedings of the Winter International Symposium on Information and Communication Technologies*, pages 1–6. Trinity College Dublin, 2004.
- [392] SurveyMonkey. Privacy Policy.  
<https://www.surveymonkey.com/mp/policy/privacy-policy/>, October 2013.
- [393] SurveyMonkey. Security Statement.  
<https://www.surveymonkey.com/mp/policy/security/>, September 2013.
- [394] Eugene Suzuki. Interview with Eugene Suzuki. Interview, September 2013.
- [395] Keiji Takeda. User identification and tracking with online device fingerprints fusion. In *Security Technology (ICCST), 2012 IEEE International Carnahan Conference on*, pages 163–167. IEEE, 2012.
- [396] TechTerms.com. RAM. <http://www.techterms.com/definition/ram>, October 2008.
- [397] TED Conferences. About TED, History. <http://www.ted.com/pages/16>, November 2013.
- [398] Bob Tedeschi. Giving Away Information, but Increasing Revenue.  
[http://www.nytimes.com/2007/04/16/technology/16ecom.html?\\_r=1&ref=technology](http://www.nytimes.com/2007/04/16/technology/16ecom.html?_r=1&ref=technology), April 2007.
- [399] Tedstaff. TED reaches its billionth video view! <http://blog.ted.com/2012/11/13/ted-reaches-its-billionth-video-view/>, November 2012.

- [400] The Telegraph. Mobile Twitter User Demographics and Behavior. <http://www.nashuatelegraph.com/digitalmarketing/digitalmarketingblog/996809-459/mobile-twitter-user-demographics-and-behaviors-revealed-in-study-by-complete-com.html>, March 2013.
- [401] Wei-Guang Teng and Hsin-Hsien Lee. Collaborative recommendation with multi-criteria ratings. *Journal of Computers*, 17(4):69–78, 2007.
- [402] Matthias Teschner, Bruno Heidelberger, Matthias Müller, Danat Pomerantes, and Markus H Gross. Optimized spatial hashing for collision detection of deformable objects. In *VMV*, volume 3, pages 47–54, 2003.
- [403] Jennifer Thom, David Millen, and Joan DiMicco. Removing gamification from an enterprise SNS. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1067–1070. ACM, 2012.
- [404] Tidylife, Inc. CreateDebate. <http://www.createdebate.com/>, November 2013.
- [405] Stefan Tilkov and Steve Vinoski. Node.js: Using Javascript to Build High-Performance Network Programs. *The Functional Web*, December 2010.
- [406] Danielle Timmermans. The impact of task complexity on information use in multi-attribute decision making. *Journal of Behavioral Decision Making*, 6(2):95–111, 1993.
- [407] Robert L Tolliver, David S Carter, Suzanne E Chapman, Phillip M Edwards, Jeanie E Fisher, Annette L Haines, Lana E Krolikowski, and Rebecca M Price. Website redesign and testing with a usability consultant: lessons learned. *OCLC Systems & Services*, 21(3):156–166, 2005.

- [408] Catalina L Toma and Jeffrey T Hancock. Self-affirmation underlies Facebook use. *Personality and Social Psychology Bulletin*, 39(3):321–331, 2013.
- [409] TopPoliticalSites.org. Top political sites. <http://toppoliticalsites.org/>, 2013.
- [410] John M Trussel and Linda M Parsons. Financial reporting factors affecting donations to charitable organizations. *Advances in Accounting*, 23:263–285, 2007.
- [411] Truthmapping. A Tool To Elevate Debate.  
<http://truthmapping.com/about.php>, November 2013.
- [412] Michail Tsikerdekis. The choice of complete anonymity versus pseudonymity for aggression online. *eMinds International Journal on Human-Computer Interaction*, 2(8):35–57, 2012.
- [413] Marlene E. Turner and Anthony R Pratkanis. Twenty-Five Years of Groupthink Theory and Research: Lessons from the Evaluation of a Theory. *Organizational Behavior and Human Decision Processes*, 73(2):105–115, 1998.
- [414] Navin Kumar Tyagi, AK Solanki, and Manoj Wadhwa. Analysis of Server Log by Web Usage Mining for Website Improvement. *IJCSI*, page 17.
- [415] Jan van Lohuizen and Robert Wayne Samohyl. Method effects and robo-polls. *Survey Practice*, 4(1), 2013.
- [416] Jordie van Rijn. The ultimate mobile email statistics overview.  
<http://www.emailmonday.com/mobile-email-usage-statistics>, 2013.

- [417] Robert Van Rooy. Being polite is a handicap: Towards a game theoretical analysis of polite linguistic behavior. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 45–58. ACM, 2003.
- [418] Amy Vaida, Wendy C Newstetter, and Elizabeth D Mynatt. When conventions collide: the tensions of instant messaging attributed. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 187–194. ACM, 2002.
- [419] Mike Vorhaus. Would you pay money to see your favorite site ad-free? *Ad Age*, pages 1–2, December 2008.
- [420] David Vronay, Marc Smith, and Steven Drucker. Alternative interfaces for chat. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 19–26. ACM, 1999.
- [421] W3C. Web content accessibility guidelines (WCAG) 2.0. <http://www.w3.org/TR/WCAG20/>, 2008.
- [422] W3C. W3C Mission. <http://www.w3.org/Consortium/mission.html>, 2012.
- [423] W3C. Web storage. <http://www.w3.org/TR/webstorage/>, July 2013.
- [424] W3C. What is HTML. [http://www.w3.org/wiki/HTML/Training/What\\_is\\_HTML%3F](http://www.w3.org/wiki/HTML/Training/What_is_HTML%3F), February 2013.
- [425] W3Techs. Usage Statistics and Market Share of Server-Side Programming Languages for Websites. [http:](http://)

[//w3techs.com/technologies/overview/programming\\_language/all](http://w3techs.com/technologies/overview/programming_language/all),  
December 2013.

- [426] W3Techs. Usage Statistics of Client-Side Programming Languages for Websites. [http://w3techs.com/technologies/overview/client\\_side\\_language/all](http://w3techs.com/technologies/overview/client_side_language/all),  
December 2013.
- [427] WebPlatform. The Basics of HTML.  
[http://docs.webplatform.org/wiki/guides/the\\_basics\\_of\\_html](http://docs.webplatform.org/wiki/guides/the_basics_of_html),  
September 2012.
- [428] WebPlatform. What is CSS?  
[http://docs.webplatform.org/wiki/tutorials/learning\\_what\\_css\\_is](http://docs.webplatform.org/wiki/tutorials/learning_what_css_is),  
September 2012.
- [429] WebPlatform. Why use CSS? [http://docs.webplatform.org/wiki/tutorials/learning\\_why\\_we\\_use\\_css](http://docs.webplatform.org/wiki/tutorials/learning_why_we_use_css),  
September 2012.
- [430] Zack Whittaker. Why do Facebook users hate change? <http://www.zdnet.com/blog/igeneration/why-do-facebook-users-hate-change/12863>,  
September 2011.
- [431] Wikipedia. Wikipedia: Database download.  
[http://en.wikipedia.org/w/index.php?title=Wikipedia:Database\\_download&oldid=584015676](http://en.wikipedia.org/w/index.php?title=Wikipedia:Database_download&oldid=584015676), December 2013.
- [432] Wikipedia. Wikipedia: Productivity of Wikipedia Authors.  
[http://en.wikipedia.org/w/index.php?title=Wikipedia:Productivity\\_of\\_Wikipedia\\_Authors&oldid=484867403](http://en.wikipedia.org/w/index.php?title=Wikipedia:Productivity_of_Wikipedia_Authors&oldid=484867403), December 2013.

- [433] Wikipedia. Wikipedia: Wikipedians.  
<http://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedians&oldid=585255487>, December 2013.
- [434] Nancy Willard. Off-campus, harmful online student speech. *Journal of School Violence*, 2(1):65–93, 2003.
- [435] Ashley Williams. Usercentered design, activitycentered design, and goaldirected design: a review of three methods for designing web applications. In *Proceedings of the 27th ACM international conference on Design of communication*, pages 1–8. ACM, 2009.
- [436] Diane K Willimack, Howard Schuman, Beth-Ellen Pennell, and James M Lepkowski. Effects of a prepaid nonmonetary incentive on response rates and response quality in a face-to-face survey. *Public Opinion Quarterly*, 59(1):78–92, 1995.
- [437] James Q Wilson and George L Kelling. Broken windows. *Atlantic Monthly*, 249(3):29–38, 1982.
- [438] Joyce M Wolburg. College students’ responses to antismoking messages: Denial, defiance, and other boomerang effects. *Journal of Consumer Affairs*, 40(2):294–323, 2006.
- [439] Project Wonderful. About Project Wonderful.  
<https://www.projectwonderful.com/advertisewithus.php>, 2013.
- [440] Stephen Worchel and Jack W Brehm. Effect of threats to attitudinal freedom as a function of agreement with the communicator. *Journal of Personality and Social Psychology*, 14(1):18, 1970.

- [441] WPI. Password standard.  
<http://www.wpi.edu/academics/CCC/Policies/passtand.html>, 2008.
- [442] Henrik Wramner. *Tracking Users on the World Wide Web*. PhD thesis, Master Thesis, Royal Institute of Technology, 2011.
- [443] Kevin B Wright. Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3), 2005.
- [444] Tatu Ylonen and Chris Lonvick. The secure shell (SSH) protocol architecture. <http://tools.ietf.org/html/rfc4251>, 2006.



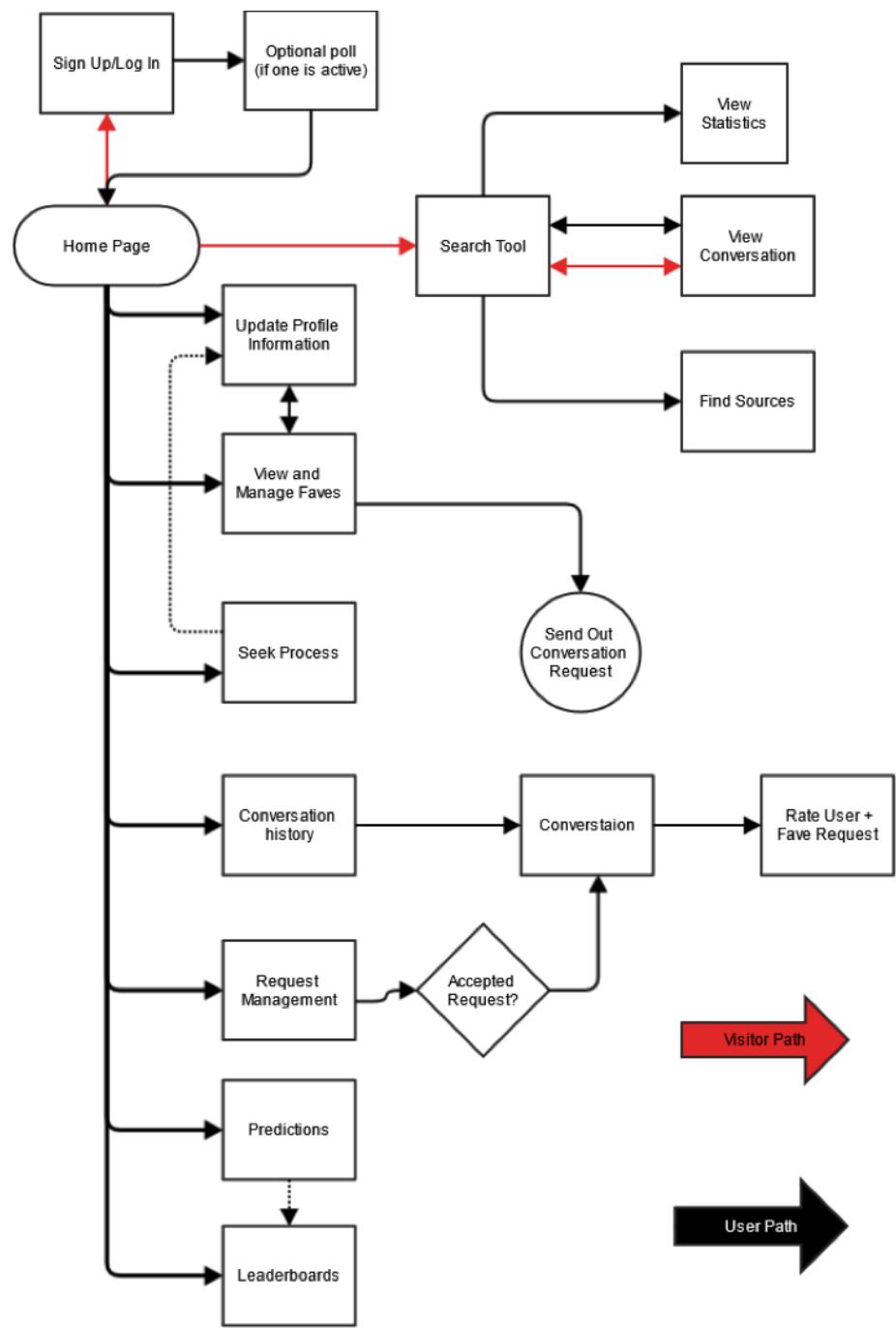
## 17 Acknowledgements

We would like to thank the following individuals for their support and contribution to the success of this project thus far.

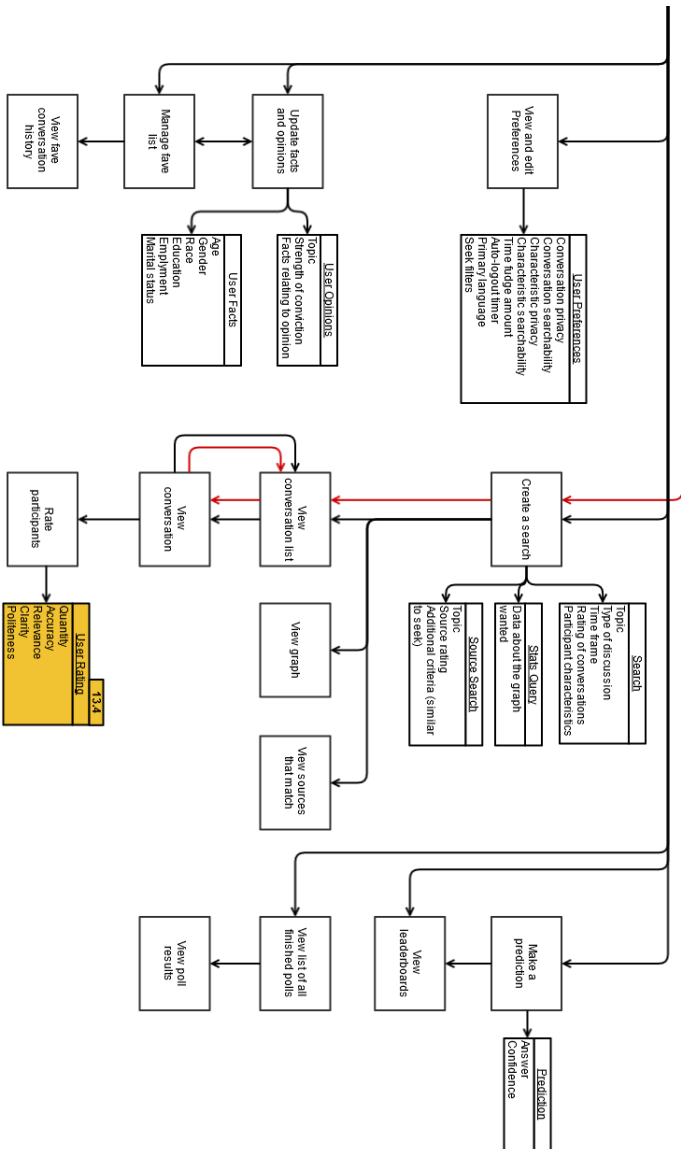
- Mr. Alan Mandel, the creator of the Civil Good concept, for sharing his thoughts, ideas, and suggestions with us iteratively over the course of this project.
- Mr. Eugene Suzuki for discussing methods of monetizing websites and the challenges of website security.
- Stuart R. Malis, Esq. for reviewing the proposed privacy policy and providing feedback.
- Ms. Ruth McKeogh and the WPI Institutional Review Board for their helpfulness and quick processing of our IRB application.
- Mr. Andreas Koeller for his LaTeX report template.
- Mr. Robert Bertsche and Mr. Peter Caruso for general advice on legal issues that Civil Good may face.
- Professor Soussan Djamasbi for advice on website user interactions and website design.

# Appendices

## A User Flow







## B Interview Transcript, Mr. Eugene Suzuki

...

Gene: So you're look for ways of monetizing on this site and?

Tushar: Yeah, mhmm?

Gene: I guess the couple of good questions that I asked Ian earlier were: is this intended to be sort of non-profit? or be for-profit? Does it not matter? because obviously given the answer to that, you have a lot of different options.

Dave: Alan wants it not-for-profit.

Gene: Not-for-profit? But not necessarily non-profit?

Tushar: Right.

Gene: So, not a 501(c), but not for profit.

Everyone: Mhmm.

Gene: So you can actually use any kind of way to make money as long as you are making money to support the operational costs.

Tushar: Right.

Gene: Great. So, that's good. All right. Something that we talked a little bit earlier too, was that uh based on the fact that uh—what was your example? You said if you had like a bisexual who was looking for—?

Ian: Yeah, if you're a female, bisexual person from the South who supports Romney and want to talk about abortion, you could do that.

Gene: So, I know that PACs would definitely pay a lot of money for that particular demographic targeting. So display advertising inventory on your website would be very useful. Only if you could achieve a certain critical mass of ad impressions per day.

Tushar: Okay.

...

Ian: And so would you say that display-retargeted display advertisement would be a viable means to support the operational costs of a site like this?

Gene: Well, I guess, your goal for monetization is to cover the operational costs. So do you have a business plan? Do you have some idea of how much the operational costs are going to be?

Tushar: We haven't gone that far in depth.

Gene: Okay, so you're trying to figure out how to run it without knowing how much it's going to cost. You're probably going to want to figure out how much it's going to cost.

Tushar: Okay.

Gene: Right? So if it turns on that you need way more or way less, then you don't want to actually become a for-profit by accident. Which is okay, I suppose. You could always figure out what to with the money. Buy beer or pizza or something. laughter\*

Gene: Okay. There's lots of different monetization models. There's always the traditional—beg— you know, ask for donations.

Ian: Wikipedia does. It seems to be working for them.

Gene: Yeah, but they have global reach. They probably have 10,000,000 impressions minimally a day.

Mike: So maybe that's not a good way to start?

Gene: Well, I don't know. If you can prove the availability of that many impressions, I'm sure that there is somebody who is willing to buy that inventory. I guarantee it. I could probably put you in touch with the person who would want to buy that inventory from AOL or Google or whoever. They would be really excited to have that business.

Ian: What about? I talked to you earlier about a recurring, crowd-sourced, paid subscription model like Subbable. Do you think that would be viable?

Gene: Of course. All of these things are viable, but to what end? How much money do you think you're going to need to maintain this site. If you have your initial upfront costs? this is a business, right? Whether it's a for-profit business or a not-for-profit business, it's still a business. You're still going to have expenses. So, let's say that you buy some shared server on Rackspace or AWS and you put up a website and it costs you 100 dollars a month to host it. That's within the ballpark, right? So, you've got the website, and you're probably going to use all open source stuff cause it's free. And all it will take to get it up and running is your time. And you'll probably need somebody to maintain the site. So you can pay software developers and do they do all that work out of the goodness of their heart?

Ian: So one question that we had was whether having multiple different methods of monetization; like having display advertisement in addition to Subbable in addition to like a store in addition to having donations. Would having all of those different methods of monetizing like the user is constantly bombarded with? with different ways that we're trying to raise money to run the site. Would that detract from the value per user. Would that actually drive people away from the site?

Gene: I guess it actually depends on how well you know your users. Look at the big enterprises that get donations. Like NPR for example—you guys know what NPR; National Public Radio. You've got all these different independent radio stations across the country that are "listener-supported" and they get donations. Most of the time they're getting the donations; the money that's actually running the whole operations is probably coming from a few big donors. So if you're looking on a graph the distribution of donors; on the high end of donations;

they're going to be a few of them, but they represent a huge volume of quantity of the total money. But the smaller donors really add up. If you look at the Obama campaign, everywhere you look during the 2012 election, it was like, you'd see some ad for money for the Obama campaign. And it was there again and again and again. It really depends on the demographic. If they know that you are actually someone who has donated in the past, you could look at it in terms of people who you know have donated in the past, there's a high likelihood that they're going to donate again. Right? So you're probably going to want to target them again and again and again via email, via display advertising, display retargeting. Whatever ways you can. Every time they visited the site, you'd want to hit them up for a little bit of money. So any mechanisms that you have to get that money is great. You'll also have to run some sort of analysis to see if there's donor exhaustion also. If you're trying to find people who are on the cusp of donating and not donating, those are probably your small donors. You can also build a good business of small donors, but you'll need a model.

Tushar: So when you say that we should target them again and again, it might be kind of hard here because we're trying to not have any personally identifiable information about users. So I don't know how that would function. If they paid you once, we would want those data sets to be totally separate.

Mike: It's a site where we collect the data set about the people, but we don't attach that to any of their personal data. And we can use that data set.

Ian: We don't even ask for e-mail addresses.

Gene: You don't ask for e-mail addresses?

Mike: That makes it a little more challenging.

Gene: So you have no value per record there, you're not going to make any money on e-mail remarketing. That's tough. That makes it a lot. Because you can make



a lot of money with e-mail. At our company, we make a lot of money from e-mails.

Tushar: Do you mean like selling e-mail or sending e-mail?

Gene: Sending e-mail. Like if you have a demographic and someone is looking for a particular demographic. You can make money by either selling the e-mail record with the demographic information or you can make money by sending the e-mail on behalf of the people who want that traffic. Or you can send it on behalf of yourself so that you can get a donation. You can always make them aware of your site if there's an opt-in. Is there an opt-in?

Mike: There was a mailing list that was supposed to be entirely separate?

Ian: Yeah, we talked about maintaining a separate mailing list that wasn't associated with user demographics and people could opt in if they wanted to receive information about Civil Good, but we didn't want to require people to provide their e-mail address.

Gene: not require but

Ian: But Alan specifically does not want there to be e-mail addresses in the signup process. I personally think that it would be a good idea to have it optional so that those who want to have that e-mail associated with their account can. I would want to.

Gene: The only way you could force someone to see your message these days is other than spending 10,000 dollars a day on display ads is e-mail. E-mail's free, basically. You're cutting yourself off from a huge revenue source by doing this. If that's what it is, that's what it is. Can you cookie them?

Ian: Uh, I highly doubt that. I don't remember if that's covered in the 60-page paper, but I don't think?

Tushar: I would say no.

Mike: I don't think so.

Ian: It becomes we difficult because we have all sorts of demographic information that they provide, but we can't connect it with any personally identifiable information.

Mike: So we're basically selling to their accounts on the site.

Gene: Sorry?

Mike: Like they have their little account on the site, and they have their demographic data which they use to query other people to talk to and who knows what else. But our assumption, I guess, for whatever methods we use is that they're on that account and they're using the site and they can see it (probably).

Tushar: So the way we do this is, say if you sign up for the site, if you say your display name would be Gene, all right cool. But that might not be your real name and that's fine. And every time you try to talk to someone—say you want to talk to a bisexual Republican who voted for Romney—if you want to talk to someone of a specific sex, you have to put in your own sex. If you want to talk to someone of a specific gender—

Gene: But how do you verify that those are all real?

Ian: You can't.

Gene: How do you prevent a terrorist organization from using your system as a means for communicating?

Ian: That's a really good question that we haven't addressed at all.

Gene: Because if you have all this anonymity?

Mike: I believe that Alan had something about that in his site where he was just saying that you sort of just set up a culture on your site such that people would almost act as antibodies against that. If they saw stuff going on, they would vote them off the site essentially.

Ian: We have a rating system that allows users to rate their partners after having

a conversation, and someone who is out to troll or mess with people is not going to get very good ratings, and people are not going to want to talk to them.

Gene: So it's not going to be moderated, but it's going to be reputation-based.

Ian: It can be moderated. I don't know whether that—that's not something that Alan addressed, but it's something that we may want to consider. There's definitely some sort of reputation system.

Mike: It's almost self-moderated. People can be literally voted off the site.

Gene: But anybody could register. So as soon as you register, you have a neutral rating, and you could post.

Tushar: Yes, if you get voted off, you could come back in easily.

Mike: I think there were some potential ways to prevent that or make that difficult at least.

Gene: How do you do that though if you're not tracking them or anything?

Dave: Based on the rating system like you could set it so that you only have people with a certain rating.

Gene: Well how do you get any rating at all if you are a new user. Right? It's a catch-22. You'll shut off all your new users if you do that.

Tushar: So yeah, we definitely haven't addressed the whole—if somebody gets voted off, they can come back illicitly and how do you stop that.

Mike: I forget if there was something.

Gene: Can you internally target people with ads without sharing it with other people?

Tushar: Yes.

Gene: So you could have sponsored advertisements for—so say if somebody likes Romney for sure, or they say that they do, you could put Romney ads, and that would be appropriate.

Tushar: Yes.

Gene: So you have a highly focused demographic population that you could sell ads to.

Ian: Yeah, a really detailed and fine-grained demographic.

Gene: But you're not giving any cookieable or pixel—you're not giving anything back to the advertiser.

Ian and Tushar: Yeah.

Gene: \*sigh\* I'd hate that. \*laughs\* Cause what you're saying is that you could put the ad out there, and someone could click on it, but it would be. I guess you could have ad groups or something like that. You could have some attribution of a particular demographic that it matched to.

Ian: What we would do is we would allow—we would have some sort of interface by which people could bid on advertisements specific to whatever demographic criteria they would like and we would then figure out how to display the advertisements, target it for the specific users that all these different companies are looking for. But the downside is that we have no actual way of proving that the demographic is being reached.

Gene: Which is very bad in terms of online marketing. Because the whole reason why people like online marketing is that you can see what the value of that advertising is. Like at our company, for example, what we try to do is we'll buy clicks from Google or whomever and everyday we get back the costs and the number of clicks and maybe the number of impressions. Those are the basic ones. On our side, we set up our ads so that they can be attributed back directly via the destination URL of an ad, we know when the user arrive that we can track their performance on our site all the way through until the very end; until they convert into a paying user. So we can basically join the ID of the arrival with the money

that was spent on that specific traffic and figure out how much we made on it and calculate the margins and profit. But you can't do that; you're basically cutting off half of that whole side. You're making the money on your side but you're not giving any data back to the sponsoring organization. They may not like that, which would actually decrease the value of your display inventory.

Tushar: How about a store? How do you feel about that? Like selling...

Ian: Like selling physical merchandise. Paraphernalia that people could purchase like oh get your Civil Good shirt or poster or something.

Gene: Yeah, that certainly works. That works for the NPRs of the world too. Sure That's a proven model. How big do you expect the site to be? How many visitors do you expect to be on the site? Do you have any idea?

Mike: It's so new. It's such a new idea that it's tough to find anything to even compare it to.

Gene: Do you have a—Is getting users part of the scope of this project as well?

Ian: Yes trying to figure out how to get the users and how to maintain long-term retention is a big question for us. After a person has visited the site, how do you get them to keep coming back.

Gene: Sure. You have to create a sticky experience.

Mike: And aside from a site store, we were thinking eventually if you could like build enough of a reputation, it could be almost a TED talk thing. Like it's all about discussion and debate like maybe you could have some really good debaters talking about stuff in an auditorium setting stuff.

...

Ian: What do you think of paid mobile access? For example, Blackboard has millions and millions on the Apple App Store, and the way it works is that the app itself is free and—first of all Blackboard is an educational tool

Gene: Yeah, I know what Blackboard is, I have a friend who is the senior guy there.

Ian: Well anyway, people have to pay an annual subscription to use the app, and I'm sure that drives in a lot of revenue. But do you think that for a site like Civil Good, would that be effective enough to drive enough revenue, or does it all boil down to the "do we have enough users"? question?

Gene: I guess you've chopped off one hand because you have no way to make money on what's really valuable about this business which is the very very focused ad targeting. I don't know? You have to figure out how you're going to get users for your site, right? If all of you had some post on your social media sites, with a one-line that's very catchy, people would come to your site. So you've got users to your site, you need to have some way to make their experience sticky. But you don't want to collect any e-mail addresses.

Ian: We don't want to force people to have to provide their e-mail addresses.

Gene: So it's very difficult to know if it's one person or ten people who are actually commenting about a particular topic.

Tushar: Well, conversations on the site are 2-way only.

Gene: 2-way?

Tushar: Yeah, so only you and I can talk together and that's it. You can view a conversation with somebody else, but a single conversation is just between two people.

Gene: So it's not like a blog or a message board.

Mike: It allows for anonymity with accountability because the other person is seeing what you're saying specifically

Gene: And you want other people to be able to see your conversation. That's what's different between this and e-mail.

Ian: And also, if desirable, it can be completely anonymous, you don't have to provide any information—even your user account when talking to a person.

Gene: So why would other people read your rants I guess?

Ian: Because if you have really high ratings, and you're known to be able to very eloquently articulate your points and you make really solid points and you have all the sources to back it up, and someone wants to follow your points, they can see other conversations you've had with other people.

Mike: So it allows people to see other people's point of views, in a world that's increasingly like filter bubbles and what-not, you know?

Gene: So unmoderated, unfiltered, discourse between two people. And you get to be a fly on the wall.

Tushar: If they so choose. If the two people agree to keep it private, they can totally do that.

Gene: The advantage here is that you're getting to talk to someone based on what their registered opinion is of a particular topic.

Ian: And the demographic of the person.

Gene: And the demographic of the person. So If I want to talk to someone who is female, bisexual and likes Romney, and doesn't like abortion, I can find that person using the site.

Tushar: As long as there is one.

Ian: As long as the person exists.

Gene: If that person exists. What if there are multiple?

Mike: Then you have options?

Ian: You create a "Seek" which is kind of like a Craigslist ad posting saying "I want to look for this person" and a person who matches that demographic can click on the Browse section for seeks that match them and see your listing and go

“Oh, this person wants to talk to me”, okay.

Gene: So what’s to prevent people from using this board to do something like hook up?

Ian: I mean, nothing.

Tushar: Nothing, but you don’t know who they are in real life unless they tell you. So I can only see your display name, but if I’m having a conversation, and you tell me where you live, well that’s your problem.

Gene: so I’ve been doing this Internet thing for 16 years, 17 years. In my experience, the Internet turns into the lowest common denominator of what you offer. I mean, I’m sure you’re going to be able to accomplish what you’re trying to accomplish, but there’s going to be a lot of subgroups of these users that will use this system for other things like solicitation or Craigslist kind of hookups given the anonymity that exists in the system. Because you’re giving the something they don’t have, which is that anonymity. Craigslist at least still has e-mail addresses which is dangerous because then law enforcement gets involved. But you’re actually providing a mechanism for law enforcement NOT to have any tracking capabilities other than potentially scanning your IP logs.

Ian: That’s all we can provide. The only thing that we have to provide is IP logs.

Gene: So there’s definitely a risk management aspect to this too that you should consider. You guys can put two and two together, if you create a completely anonymous site where two people can talk completely privately but you can find people of particular proclivities or alignments or whatever, you’re going to get these kinds of things happening. It’s just going to happen. Which may be okay too because maybe that’s what you want. There’s some liability associated with it too.

Dave: Alan mentioned something about a reporting system.

Ian: There is a—if you actually look at the mockups I made, at the end of a



conversation, in addition to being able to rate the user, there's an "Object" button which presumably if you click on that, you're reporting the user for some sort of inappropriate behavior.

Dave: Like either revealing their real name or flaming

Gene: But they can always just register again, though. That's the back door. You can burn your account and start a new one.

Dave: Well maybe if we collect IP addresses, that can help prevent that. Because we're going to have them anyway.

Mike: I think that was the idea; if you have an IP, we can IP ban them, almost like 4chan.

Dave: The problem is IP spoofing.

Gene: And also have legitimate users behind a NAT firewall.

Tushar: Well, so I have a question for you both, so sites that collect e-mails. How is that preventing you—if a site bans an e-mail for inappropriate behavior, what's to stop me from registering a new free e-mail account and signing up. So basically you're just adding more friction to the process where if I'm really dedicated about getting back on the site, I just have to create a new email, right?

Gene: That's right.

Tushar: And that extra step of friction is what we are missing.

Mike: I think there were some other?

Ian: Or you could be like me; I have a? I have my own domain, and I have just a catch-all email address, so if I want to send an email, I can send it from anything@ianonavy.com, and anything sent to my domain just gets sent to one catch-all email address.

Gene: This is a function of moderation, right? Somebody has to do it, right, or else you won't have a viable site.

Ian: Right.

Gene: That's somebody's time. Not something that's very much time, but it's a job.

Ian: Well, it's I mean, in other sites and other communities, there has been much success in volunteer-based moderation. Like reddit, all of the people who actually manage these subreddits and enforce all the rules are entirely volunteer-based. It seems to work out pretty well for them. Even Wikipedia, all of the moderation on Wikipedia is almost entirely volunteers. It's really impressive.

Mike: It's sort of a culture that you build up on your site.

Gene: Sure.

Mike: And that's the idea with our site. You could hope for some self-moderation, hopefully. If you can build that positive culture.

Gene: So monetization, you have donations, you have store, uh, you have display advertising but you're not giving the feedback of the efficacy of that advertising back to your buyers, you can't really do email retargeting. You can do opt-in email.

Ian: Well, we can't tell them, like, we can't give them cookie information, but we can tell them that a person of a specific demographic clicked their ad.

Mike: That's true.

Gene: Oh, that's true. That would be okay.

Tushar: But again like you said, that's what the person says they believe in. So I don't know if that drives down the value.

Gene: Well, what it basically means is that you'd have to have your own ad server, and you'd have to display the ads from your side as opposed to from some other side which by virtue of having the ad displayed could implicitly pixel you. So if you took them out of the entire equation and you got the ad from them, you would

become the ad server also.

Ian: And it's the only way to do it because if we have other people do the advertising, we don't have any control over what ad actually gets displayed.

Gene: So, it's all about critical mass. Can you get a critical mass? If you can get a critical mass, then you can actually start get paid quite a bit more. So there are three primary models of Internet advertising. There's the cost per click model where you charge people based on a click. There's the cost per impression, or actually the cost per thousand impressions (they call it CPM, cost per mil, they call it). So 1,000 impressions might cost you 10 dollars, which means each impression is one penny. So that's another model. And then there's the CPA model, which is the cost per acquisition model. The cost per acquisition is somebody comes to your site and you get paid if somebody turns into a paying user. And then there's the cost per lead model (CPL) where you get to the destination site and you provide enough information where you're a viable lead. On our site, we have a landing page and there's a ZIP code field and then some question. And you fill out the ZIP code and the question, and you become a lead on our site. So we buy that way also. So there's four primary models: CPC, CPM, CPA and CPL. Depending on your demographic and depending on what your money requirements are, you could probably figure out how to—what you need to charge in order for it to back out. And also if you want to run an auction or just trying to have somebody try to sell those ads directly to anyone who wants that particular demographic. If you need a business development person, then you're probably going to want to spend some money on that person also.

Ian: Would you say that an auction model is better than having people make partnerships for individual demographics?

Gene: Only if you have enough participants in the auction. If you don't then?

Ian: If you only have one person bidding on a specific demographic.

## **C Interview Transcript, Professor Soussan**

### **Djamasbi**

Prof. Djamasbi: So, generally speaking, the best practice these days is that you don't think immediately "What is the interface?" You have to first come up with "What is the experience that I want my users to have on this website?" And of course that depends a lot on "what is the scope?" What you want the users to do, how you want them to access the information, what is their role, right? So in some cases you can even storyboard how your user interacts with this thing, and then once you have that, then you start thinking about "ok, these are the experiences that my users want to have" for example, the things that we're talking about, like number one thing, and there will be tons of different things, but one of the most important things, anonymity is very important. Right? Obviously, once you get to a more refined version you're talking about what types of users go there, what would be their preferences, what is their method of interaction, what they usually use to exchange information, the platform they're using, and based on that you can design the experience, right? So once you have that, then you say, ok, what is the best way to implement this experience? And this is very important because when you are both the designer and the coder, you just start thinking code, right? You just can't separate them, it's like "Oh, ok, so do I have a drop-down menu, or do I want the user to type in, or click a button?" Or, you know what I mean, because you're thinking also "Oh, I know that with this platform I'll get this kind of functionality and I can develop it this way" But it is very important that you focus on the experience, right? So, having said that, now, the flow: So, the users

enter the site, and sign up. Right? And after the user signs up the user has a username and two passwords, right? And can log in with the username and a password, right? And once they enter, somehow they can communicate and it's just very much like games, like in games, you have a nickname, so you kind of have an anonymity if you are playing with an amount of other people. But this is more in a webpage. So you understand what I'm trying to tell you, and I know this is ahead of the... but like If you guys came to me at the beginning of A-term I would have made you guys to make storyboards and stuff like that.

...

Michael Perrone: anyways, account setup: two passwords, and then it asks you for a LOT of information, and the general idea behind this is to actually raise barriers to entry, to make it less worthwhile for anyone who would want to like, troll on the website, or just -

Prof. Djasmasbi: And that's a good way, right? Because exactly that's what - if you want to do something with a lot of people, you would recommend not to do it.

...

Prof. Djasmasbi: But they can lie about it (their information) too, okay?

...

Prof. Djasmasbi: And the other thing is that if they don't lie about it, what if somebody hacks into this information? Are there any liabilities, or - you don't ask them for - there's no identifiable - well, neighborhood, ok, type of residence, type of occupation... you know what I mean? I think that's - that's what your company or sponsor has thought about, probably. But yeah, they could lie about it -

Michael Perrone: And then that would lower the barrier to entry.

Prof. Djasmasbi: That would lower the barrier, but from a design standpoint, if you want to discourage somebody, that would be something that you would do.

I'm pretty sure you have ... you have something on your homepage that explains what is the purpose of this.

...

Prof. Djamasbi: Because, you know, you can discourage people, and you can encourage people, and you want to do this at the same time. But, on your homepage, you want to attract people who you want to attract. So you have to have a message that is very attractive and that is very understandable to people who could be contributors to this.

Michael Perrone: That sounds like... The sign-up page that was originally developed ... but the idea is, he has a video explaining what the site is, and he goes into a bunch of detail, and I guess the people who would go look through that and be interested would be the types who would go into detail and care about the debates.

Prof. Djamasbi: That's very good. Another thing you want to do is the SEO, so that people who are looking for this, they will end up on that site. Ok, so that's good. So: this is not a bad design, for the purpose that you're telling me

...

Prof. Djamasbi: Um, yes, you may, you know, some people, like, I generally don't like when people start asking me questions, "Um..." I'm thinking, "Why? Why do you want all this information about me?"

...

Prof. Djamasbi: And, then again, there is no way you can please everybody, so some people, you're going to lose. Another thing is that you may want to also check about the contrasting colors on the website to make sure that people with various disabilities can access this website, and make sure that it's also friendly to screen-readers, for blind or low-vision people. Now, a lot of people who are not

blind, they still use screen-readers, because of you know, their poor vision. And a lot of people who are dyslexic also use screen-readers, so I think for a forum like this, the premise is that you would want to include everybody.

Michael Perrone: As many people as possible - yeah.

Prof. Djamasbi: And it would be, counter -philosophy of the website to exclude those people.

...

Prof. Djamasbi: So, so the contrast, you want to make this better, but there are guidelines actually, for the color contrast on all of these things. The screens - are these all going to be different screens, or are these going to be one screen?

Michael Perrone: I think they're one screen. (talking about the home page in the Civil Good design document)

Prof. Djamasbi: Allright. Because it looks very cluttered to me.

...

Prof. Djamasbi: Another thing that often helps and actually if you give the user a choice and they can decide, that is the optimal thing, and then actually have a statistic to show how many of your users choose that option or don't choose that option. And that supports the functionality or not - if you have to choose, right? That particular thing, but I think that would be one way to go. And, that could be like, conditions like you say, I want to find a democrat: Age range, gender, what type of conversation do you want to have: synchronous, asynchronous, you know what I mean. ...

Prof. Djamasbi: It's very textual... I mean, visually, what you try to do - but you say that you're also trying to discourage anybody from, you know, engaging this one. So if that's your overall thing, somebody has to make the time - invest the time in it to do this. And if they do, then they're pretty much familiar with "This

is about that, this is about that”, you know, and they can... familiarity is going to help, especially if you’re not changing the look every, I don’t know, two months or something... So that shouldn’t be a problem, however it goes against the visual design...

Michael Perrone: Rules?

Prof. Djamasbi: Rules and stuff like that. My guess is that a lot of people who are attracted to this, are probably people who are very textual because they type, they write...

Michael Perrone: So it might create a bias in that respect? That’s one worry...

Prof. Djamasbi: Well, there is no way that you can please everybody. You can try to cast the net as much as you can, but if you want it to be anonymous, text is the way to go. If you want it to be an audio, would that be a mode of communication that somebody can leave, you know? So, this is step one: you want to design something that is data-driven, so if this is your first step, you do it textual, and then you can collect information, just like the surveys that they have, “If they were going to expand this, what would be your preferences?” You know, “Would you want audio?”

Michael Perrone: So, build that into the...

Prof. Djamasbi: Right, build into that... obviously video is not going to be part of it because it’s - you will be recognized, right? And then you will see what are the costs of that, because there are a lot of users with a lot of various preferences and then the cost becomes a determining factor a lot of times. So I have worked with companies that they say, well when we survey, customers always say “oh we want preferences, we want to be able to personalize things” but they don’t use those personalized features. So you know, you could put a lot of effort into designing features and then people don’t use it, or use only a few, but in your case, you



could just build one step and then ask them. For example, you may find out if you have so many options, they choose only a few.

Michael Perrone: and then those are the ones to elaborate on, and the others, you would get rid of, phase out like...

Prof. Djasmasbi: Well at a minimum you'd prioritize the order, because we know from eye-tracking studies that you look from top to bottom and you pay attention on the top

Michael Perrone: yeah, and to the left

Prof. Djasmasbi: And to the left. So you lose interest, so you could prioritize them, based on that. And at some point it could be even... make it like a sandbox: give people the options to organize stuff the way they want to do it. If you give that option to people, then you make it more palatable and you may attract some people who can really ease in. But this is, I guess, after you get 'em and they sign up, then, having the options to move things around, that may not be a bad idea. All of this has to be data-driven, you know? Because I have seen so many times, we assume things, we go there, and our clientele...

Michael Perrone: Doesn't like it.

Prof. Djasmasbi: Behaves completely differently. So, sometimes you have personas, say "Ok, my clientele has these preferences and works, or does, this way." And a lot of times it's true and a lot of times personas don't work the way they're supposed to. So data is the best thing to do.

Michael Perrone: So we can finagle as much of the designing as we want but it's best to just test it.

Prof. Djasmasbi: I think this is your best bet, because if you want to make informed decisions like, get some people who are likely to - I mean who you think would be interested to try and test it and give you feedback; what they think,

right? Or if you do it with...

Michael Perrone: We have a small survey thing that went out, so that's perfect.

Prof. Djasmasbi: Yeah, exactly. So if you're doing it only with student population...

Michael Perrone: It might not be generalizable?

Prof. Djasmasbi: No: students are Generation Y, the younger generation, has a completely different worldview in terms of anonymity, and you know, technology use, and preferences... which may bias your population. So you can go and um... I don't know where would be a good place... Like, I would say you would go to a political party and see how many people... you know, like, Republicans and Democrats, I think this is the most controversial, I mean in terms of like, they don't talk to each other, and see if they would be interested in something like that. But then again, you're not limited to - this is not limited to political debate, right? ...

Prof. Djasmasbi: Yes... And I don't understand what is your definition of this population, it's "good enough for debating" I don't know what you constitute as...

Michael Perrone: Well I mean, on the web, the idea is... Civil good tries to address the problem that on the web people will try to discuss something in a group and then their discussion will get "derailed" and they'll move on to some other irrelevant topic, or someone will "troll" - essentially try to annoy or pester people just for it's own sake and that ruins the depth of the conversation.

Prof. Djasmasbi: Ok so, a good, a typical user or an ideal user of this site would be...

Michael Perrone: Would be someone who actually uses the site as it's intended.

Prof. Djasmasbi: Which is carry on a debate without derailing it, or dominating the conversation or something like that?

Michael Perrone: Yeah.

Prof. Djasmasbi: So... That's a pretty ambitious goal without monitoring it. Is it? I mean, how do you, how can you know, well I mean, think about when two people talk, right? Or when you're listening to the debates, are you saying that, is the premise that if it's anonymous it's more likely to be civil? You know what I mean \*laughter\* I mean, this is so characteristic of people - when I see the debates, something very controversial - of course we're going to have debates that are not so emotional - but if you get the debates that are very controversial you have very strong feelings, and it's just like one is persuading the other to agree upon that. So... derailing it? I understand that the premise is to have the ideal user that looks like this, but what's your evidence that you have a lot of users like that? So you're building something based on the premise that - what happens if your ideal users are very few? So what's going to happen?

Michael Perrone: Yeah, you're saying the site's going to be small.

Prof. Djasmasbi: The site is going to be small, or I go in as an ideal user and I'll see, "You know what? This is not working, I'm not gonna go there." So you're gonna...

Michael Perrone: So even if you are an ideal user

Prof. Djasmasbi: Even if you are an ideal user, and you'll see...

Michael Perrone: That there's not many people there so it's not as useful.

...

Prof. Djasmasbi: Right. So having two, I think, is more manageable, but... well what we're discussing are at very many different levels. So we're discussing design, we're discussing philosophy, we're discussing policies, right? And that's the way it usually is because one affects the others. So yeah, following with this we enter our username and we get in, answer questions, and if we're still not deterred from answering all those questions, right? Then you create an account and you log in,

right? Ok, Is there a poll running that users qualify for, or is interested in? That's optional, ok.

Michael Perrone: We made that optional but it was, if I remember correctly, mandatory before... So is that a good decision?

Prof. Djasmasbi: I think optional is a good idea - If you want to - this is my personal opinion though, not as a designer - If this is such a - and democratic is the only word I can think of, but I don't mean, like a democra-

Michael Perrone: Yeah.

Prof. Djasmasbi: You want to have a platform that people can come in - you don't want to force them to do anything - right? So...

Michael Perrone: So in our big re-design of the thing we moved (most of) the big question set that you would have had at the beginning, and we decided on a model where if you do a seek on a certain type of conversation, we will ask you what your opinion belief is on the topic, and we'll ask like, "you searched for these qualities of other people, like age, general living area... can we have these about you?" So instead of getting that all at the beginning, it's a database that fills in as people use it.

Prof. Djasmasbi: I think it should be optional - well - then again this is my "as a user" preference: because it seems like some people are introverts and some people are extroverts, right? Some people don't mind to be the debaters, and some people would just want to listen to the debates. So you want to attract both, right? So if you give them an option, like can we, just listen to a debate or can we participate in a debate or something like that and we need this information from them. That would be good. So you're creating like a match.com, but for debate And I guess even in match.com you can just view other people's profiles, or, or... I don't know, I've never been to the site before, but I think... Anyway, but you're giving the

choice to people, to be part of those who debate, and then, do you want to have an incentive to... Is it something that you want to encourage people to debate, or do you want them to...?

Michael Perrone: One thing we were considering was gamification of the process: some sort of score for being able to change people's opinions and who knows what else - for taking more polls, for making predictions that turn out to be correct, et cetera.

Prof. Djasmasbi: Yeah, these are all good things, or... It depends of what you're after. So if you're specifically interested in improving the quality of the debates, then work with parameters that specifically encourages that, such as "how many viewers does it have?" or like an amazon rating of "what people think about this"... You know or hotel.com - you know what I mean? Something like that and then you have very specific criteria, for example, it's not divergent, it's respectful, it's... You see what I mean? Because you could have a very controversial debate and people like to watch two people tearing each other apart, you know what I mean? And then it may get a higher rating but not because it's a civil debate, it's because it's just like a show: so think about what are the criterias and then maybe people can do that. Gamification is good, but I think for something like this you want to see how does it support the premise - or... What does the site offer?

You're building a community: Remember that. Ok? If you're building a community, trust is really important. (and) What you get out of that is really important. So, if it is too gamey, then it may lose its attraction to some people... So there has to be a balance of the two, and you can start with things - like you can poll people, as I said, you don't have to...

Michael Perrone: You don't have to start out with everything.

Prof. Djasmasbi: Exactly. So if you want to do something for sure, it's data-driven:

as you go along, you add stuff. And then your community then, is built upon the community preferences. You can even have polls about policing it, policing like “how should the debates be run?” Right? So you don’t have to make all of these decisions beforehand. Then one of your challenges will be, where do I find, and how do I attract, this group of people to build this community. Such that you have some few first building-blocks and the rest goes from there. And you may actually have to start. You may have to go get some, I mean, this company can go and find two people, start a debate, let’s have an example, you know? And then other people listen to it and then you get their comment and you build it upon that, ok? That’s the content. So for the website flow, you can get - you can test people - you see what I mean? - and see how it goes. Of course that, one you have those initial designs then you want to, again get the information from the people who are actually a part of this community or would like to be part of this community.

...

Prof. Djasmasbi: Like, strength of conviction - is that what your company asked or you guys came up with that?

Michael Perrone: Oh, um, that was one of the parameters that Alan Mandel wanted to collect. You’ve got - when you ask about an opinion someone has you have a little - in his model he had a little slidy-bar that you move back and forth to rate how strongly you feel about...

Prof. Djasmasbi: This particular thing. I think that’s a very good thing. Ok... View list of all seeks user fits, ok. Based on that... details about the seek, decline, accept... ok that’s good. Accepted seeks, view details of others, that’s fine.

Michael Perrone: Pretty straightforward?

Prof. Djasmasbi: Yeah it’s very straightforward

...

Michael Perrone: And then also, they share these things in common with you, et cetera et cetera - try to be able to bridge the gap, whatever it may be. You know? And generally just getting people calmed down and maybe feed them a fun fact or two, something about the discussion topic that they're looking at...

Prof. Djamasbi: I think that's good, that's a good idea to inform the user. Ok Rate the user...

Michael Perrone: That part, that comes after the conversation finishes... Send a fave request if they enjoyed the conversation, et cetera.

Prof. Djamasbi: Uh huh, and here the user is the debater, not necessarily - people who just watch, they don't get any of these things. Or can people say I want to watch this conversation? So they can select all of those things but they can't give any ideas to the... or their preference is I want to give some information to the debate but I don't want to rate.

Prof. Djamasbi: ... I think that's what I'm saying, like, on some of these, I don't know, do you have some rules and regulations for example, don't use profanity, like, I don't know - Is this being monitored?

Michael Perrone: Sorry I forgot to mention this earlier - the ratings system is used such that, I mean, it's gamified so people who get really good ratings are encouraged to keep their account doing well, and stuff like that... And they get encouragement and whatever else, and then if there's someone who signed up to troll on the site and stuff like that they'll get low ratings and eventually fall out of ...

...

Michael Perrone: Yeah, so the idea is, hopefully in the long run you'll have a whole community of really good debaters who all have good scores, and then if someone

comes on who constantly tries to troll, they'll slowly fall out...

Prof. Djasmasbi: That makes sense, yeah.

Michael Perrone: And then, it can actually, if you get a low enough score, it can kick you off... So it's self-moderated in a way, amongst the community in there.

Prof. Djasmasbi: You may have like somebody who's really mean and who rates the other person really badly

...

Prof. Djasmasbi: Well you could have like an incubation period, let's say somebody gets a really bad vote, you don't necessarily, let's say you don't... I don't know I'm just thinking this... so two people, they get rating, but you have like a buffer, that... like let's say you have at least five votes. so that you get a rating, and in this way we'll allow that you get an average. So if this person repeatedly gets bad voting, then you get... His or her score comes down, right? But if it's only a one-time thing, the score comes down but it's not as bad. Right?

Michael Perrone: So as they sign up initially they get a buffer...

Prof. Djasmasbi: You get a buffer, and then so what if your first five debates, for example, you don't get any scores, and after that you have a window, right? So somehow you want to be forgiving to people who - and you might want to think about that, you know... You're right though, if this person debates for a long time, so average score, even if it gets bad ratings eventually goes up, right? And so then you have a history. That may not be a bad idea - especially to encourage people that "Oh what if I get a bad score" so they don't start... And you could also make that a preference, and say, "would you like - if you're starting new - How big do you want your buffer to be?" Somebody may say zero, and that's their choice. Right? Or uh... And I don't know how many "bells and whistles" you want to build, these are all things... But I think something like five, or you know... I don't



know... yeah! Five. depending on... Some people, may do one debate, and they may not go back, and they don't have any rating, so if they care about their rating, they will have to have at least five to get a rating up, so it could be part of the gamification too! Now the question is it five, is it three, is it two?

Michael Perrone: That's something to play around with.

Prof. Djasmasbi: To play around with, you know, what would be incentive? Some people are more conservative, some people are not, but as you build it, these are the information - you will get a lot of behavioral information.

...

Prof. Djasmasbi: Right, Yeah. So I think that's a good thing. And then you have to also remember that for this thing to be successful it has to have a critical mass, and it has to continue to kind of mature. So somehow the users that get there, they have to have an understanding that they're helping to build this up, and that is like crowdsourcing, right? And I think it's... hopefully it goes through a self-selective process, and then people who give up and they don't go back, you know... people stay with you who care enough about building this community, you know. And that's exactly the kind of people that you want, right?...

...

Prof. Djasmasbi: Alright, ok. Everything is viewable to everybody.

Michael Perrone: Um... Oh, up at the top here, visitors can view some things...

And users can actually do all these conversations and seeks and stuff, and favorites and have opinions and stuff, and then just viewers who aren't users can search the database maybe and view past conversations that were highly rated or whatever...

And then perhaps, add another rating in, uh I don't know, that's just this little stub on the end there. But it's all... Stuff we could do.

Prof. Djasmasbi: Because you want to convert visitors to users, so therefore you

give them more privilege if they're users, ok.

...

Michael Perrone: Yeah, that's what I was... Although we did make some mockups that I'd like to show you...

Prof. Djasasbi: Right but your mock... You would benefit from it when you want to refine, after you do the survey after you build it after... you know, it's like the second generation you may benefit from it, but at this point, you have bigger problems.

Michael Perrone: So early on you want the flow charts...

...

## **D Interview Transcript, Mr. Rob Bertsche and Mr. Peter Caruso**

...

Rob: Well, you're not my client. I haven't taken you as my client, I haven't put you through my conflicts as a client. I thought this was an informational interview...

Steven: Well it really is, it's gonna be just information about what we need to be aware of when the implementation does happen.

Rob: OK so we'll be happy to talk to you broadly about what a site needs to look at generally and that sort of thing but we're not giving you any specific advice relating to your situation and to do that we'd have to know a whole lot more, spend a whole lot more time with you then frankly you have available to us. OK?

Steven: Yep, that's fine.

Rob: But this is sort of generic conversation about these issues.

Steven: Yeah.

...

Steven: So users will be asked for demographic information and also their opinions on political topics, not all at once and not at sign up, but gradually over time.

Due to the need to comply with all state law which can be very different we're planning to just treat everything that we can as Personally Identifiable

Information to comply with any definition that the states might come up with. We plan to encrypt as much data as is reasonably possible. We also plan to have a privacy policy and keep that updated and posted on the website. And we also plan to comply with the Federal Trade Commission's Fair Information Principles. Does that all seem sufficient to you?

Peter: It's a good start, that's a good broad start. What you'll need to do in each case and there is always a danger in creating sort of a default definition of all information, and it sounds like its a good idea to begin with, but if there is some sort of breach for Personally Identifiable Information as it's defined in all of the different state statutes there are definitely breach notification provisions that you need to follow and if your policy is indicating to your consumer, and this goes for any policy on any site of this sort, you're telling your consumer that we will be protecting any and all of your information including demographic information and political views, which are not to my knowledge in any definition of personally identifiable information in the 50 states. If that is the case and there's a breach of demographic or political information that you've stored on the website, you may have suddenly put yourself in the position that you need to do more than you actually need to as far as the breach notification, but having said that, the default to treat everything in an encrypted fashion and follow, you know, a good set of rules is to follow the most restrictive definition of Personally Identifiable

Information and the regulations propagated by that. If you do that and treat all the information as such then the likelihood that you're going to run afoul is greatly diminished. A lot of that is going to have to do with the type of information collected, the timing of information, when it's collected, and how conspicuous and transparent you are about how that information is going to be used. So, it's not...what you've listed is a pretty good list of regulations to follow.

Rob: So when I look up that I want to talk to a rich person who likes using guns to shoot innocent people I'm not going to find out the name of that person.

Steven: Correct

Rob: Or any other...

Peter: Will you find out the username and/or an email address of that person.

Steven: You'll find out neither initially. At the end of the conversation you have the option to ask the user if they want to become a sort of friend, and you have a friends list. And if they accept then you'll see only their user name and their previous conversations, but otherwise you wouldn't see either of those.

Peter: OK, so if I'm a silly naive consumer and I've put PeterCaruso as my user name and I accept a friend at the end of the day. And Rob, who's smarter than I am has a more creative user name, couldn't identify him, and I friend him, and Rob finds out that his partner is a crazy gun wielding rich guy, he'll be able to know that because he'll match up my name, correct?

Steven: Correct, but that would only happen if you accepted the friend request.

Peter: Right, but I think it's also at that point in time there would have to be some disclosure that the user name can be shared after accepting the friend request and whether or not you need to put that as an opt-in, an affirmative opt-in scenario, when you decide that you're going to become a friend is something that you'd probably want to look at.

Steven: Yeah, we already plan to notify users of exactly what being a friend entails. Where it's really just you can see their user name, their conversation histories, and you can also start a new conversation with them directly, so you don't have to go through the whole search process again.

Peter: OK, and when are you, when would you be disclosing that type of information? When you sign up initially for the site or at the time that I decide that, Rob wants to be my friend?

Steven: It would probably be at the time that the friend request info-box pops up, basically, which would be at the end of the conversation. It would say "do you want to be a friend of this user?" and what does that actually mean.

Peter: OK, that's good. The general rule is that you want to make sure you put your disclosure right at the point of the selection of whatever option might result in the disclosure.

...

Steven: OK. So you talked about breach notifications which is something I hadn't thought of at all. We're currently not planning to ask users for email addresses.

So, if we don't have their email address and a breach happens, what are we required to do? How would we notify them if we don't have their email address?

Peter: Well, that makes it a little bit more difficult. But you're under an obligation to notify, to use good faith efforts to notify any one of your users that has been subjected to a potential breach so that they can take action, and depending on what state it is, some states require that you disclose the nature of the breach, and other states actually prohibit you from specifically disclosing the nature of the breach. It's not a one-size-fits-all, literally you could be sending out 15 to 25 different breach notifications if you have 15 to 25 different states involved.

Rob: But they don't even know where to send it to, so would they post something?

Peter: Well I mean, you'd have to, you could post it on the website, you could try to track people down by their computer identification number, their tag numbers. But just because you haven't collected enough of the information in order to notify people doesn't necessarily leave you off the hook. Now having said what I just said, if you're not collecting information that can identify someone, meaning their first name, their last name, their first initial, their last name in conjunction with some sort of bank account information or other information that might fall under the definition of Personally Identifiable Information, you may or may not need to worry about that because that's the definition of Personally Identifiable Information. And email is not one of the components of Personally Identifiable Information, but I'm sort of curious actually as to why you wouldn't want to collect any of that type of information.

Steven: Well, because even if an email address isn't considered Personally Identifiable Information it often can be used to identify individuals. And the founder and sponsor of this project, one of his major concerns is user privacy. And so if there was a breach, he'd want to make sure that the damage is as limited as possible considering the amount of information you might be getting. And so this way if you know someone's email address, even if you don't know their name, and you had access to the database, you could still identify them and this is just one less way that that would be possible.

Peter: It's one less way that we can do it but if something ever came down to it there would be a defamatory comment out there, there are mechanisms, there's a process, but there are mechanisms if you're a successful litigant by which you could go in and demand from the site the names of the posters and simply all they have to do is look up whether or not there's an email address or not, they'll just simply look up the identification tag numbers for the computer.

...

Steven: The next topic that I looked into was intellectual property issues because copyright violation is a very big issue in this day and age and it's certainly possible that someone could post a link to copyrighted material in a conversation to another user. And conversations, once they're done, they might also be publicly viewable, it's one of the options you can choose, if your conversation will be public or private...

Rob: It's not going to be a problem. If they post a link it's not going to be a problem.

Steven: Right, I was about to say that we plan to follow the "notice and take down" requirement of the DMCA and as long as we're doing that is there really an issue?

Peter: I guess my question is is there going to be... copyrighted information, like Rob said, that's contained in the link, posting a link is not going to violate any copyright so we can sort of put that to one side.

Rob: There wouldn't be anything to take down.

Peter: But if somebody decided to list whole cloth a conversation that they had or, god knows what, material that was generated by a third party from a third party site and just clip it in to their conversation as their own or somehow otherwise re-purpose the third party's copyrighted material then yeah you could potentially have a problem. Now taking advantage of the DMCA, there are a couple things that you need to be aware of. One of the things you have to look at is how much control the site is having over the content that is solicited, that is published, are you curating it, are you taking things down, are you editing it, etcetera. It's really out there to prevent a service provider, an internet service provider, from becoming liable for the actions of third parties of which they have no control.

Steven: Yeah we don't plan to edit anything and we'll take down stuff if we're

legally required to, like as you said, if someone just posts a copyrighted text verbatim and we get a take down request we would probably take down that conversation.

Peter: Well that's good, and there have been some publications that have gotten themselves in trouble even though they have a DMCA policy that they follow and they're protected generally under the DMCA, if they make a statement to a consumer that they're going to do X and they unreasonably delay or just flat out don't follow through on doing X whether it's inadvertent or whether its purposeful, that DMCA protection can be slightly eroded if not eliminated.

Steven: OK, that's good to know. So it would need to be a quick response time.

Peter: Correct.

Steven: OK. I also looked into trademark violations but I don't think that would really be a big issue. Do you think that that would be something that we would need to be aware of and watching for or...?

Peter: I don't think you're going to have a problem from a trademark violation if you're just concerned about your third party conversations that are going to be used, unless somebody is using a trademark for their user name, they're using a trademark to identify themselves and they're posting about that company. So if you have Company ABC that happens to trade in pharmaceuticals and there's a conversation about pharmaceuticals and somebody's user name is ABC which is the identical name of the pharmaceutical company and starts saying disparaging things about a competitor, or giving out information that purports to be information about the product that ABC is actually selling you might have a problem because ABC might then call you and say that's not us, that's somebody who is actually posting under our trademark, get rid of it because it's diluting our brand and it's causing market confusion.



Steven: And in that case is just having a similar notice and take down procedure sufficient?

Peter: It is, but you have to make sure that you act on it because a lot of times people don't do that. You can have a policy there. A lot of social media sites, Twitter being one of them, have their own trademark dispute policies by which you can submit a dispute via Twitter that they will handle internally and deal with it from there, so that's not a bad place to look.

Steven: Alright, thank you. One thing you mentioned a little while ago was defamation, and that's certainly something that we're worried about as well. I'm aware of section 230 of the Communications Decency Act...

Rob: What are you worried about?

Steven: If someone comes on and during a conversation posts defaming comments and the conversation is publicly viewable once it's finished, could we be held liable and what do we have to do to protect ourselves?

Rob: No.

Steven: Just straight up no, we don't have to do anything?

Peter: Well you wouldn't be held liable for the defamatory comments that's up there because you didn't make it, you also have the CDA, you mentioned you looked at the CDA correct?

Steven: Yes.

Peter: The CDA is going to give you that ticket of protection, in fact if someone feels that they've been defamed they can try to go after the person who's done the defaming and made the defamatory comment. And the only way you would be involved in that would be if they asked you to give up the name or the identification of the person who made the statement, and if you said, generally you would say no until there's been a subpoena issued and then they would go and

issue a subpoena and then you would simply give up whatever information you have about the poster of that content but you wouldn't be held liable for that at all. Unless of course you exercise some sort of editorial discretion and sort of put your imprint on that statement, but it would have to be a fairly substantial effort.

Rob: As a business matter you may want to be responsive to take down requests about it, where somebody's defaming somebody because it doesn't reflect well on your brand, but it doesn't create legal liability. Like Peter points out you may also want to make sure your terms and conditions leave you wiggle room if somebody asks you to disclose, you're subpoenaed for example, to disclose the identity of an otherwise pseudonymous commenter, you don't want to have terms and conditions in which you promise people you will never do that.

Steven: Right, that actually leads into the next section, which is information requests. And there's really there's two types of those that I've found, there's a subpoena, like you mentioned, that someone could issue, and there's also the whole NSA scandal that's being revealed. It's possible that the NSA may want access to Civil Good's database considering all the personal data that Civil Good may have and so that's something I researched as well. But let's start with subpoenas. If Civil Good receives a subpoena what options do we have?

Peter: Depends what the subpoena says, that's a typical lawyer answer for you. If its a subpoena for certain information it depends on what the information is, if they're subpoenaing you because there's a defamatory comment up there, and you receive a subpoena from a court, a duly authorized subpoena, then you give it to a lawyer and we decide on whether or not we're going to fight it or comply with it.

Rob: The base rule is that you have the same obligation as anybody has to respond to a subpoena. You can always challenge a subpoena in court as Peter says, so that's something in consultation with a lawyer you would decide. You

would decide, I guess, first on your own whether as a business matter you want to challenge the subpoena. You, for whatever business reason, you think it's not in your interest to be giving over that kind of information, and that business reason could include that you think it would be sort of a violation of the trust you've developed with your community. And then if you decide that's what you want to do though, now you've got to, because the baseline rule is that you have to comply with a lawful subpoena, you've got to be able to take some legal action, it could be as simple as a letter of objection or it could be something requiring court intervention where you essentially create a mechanism whereby you either don't have to respond right away to the subpoena or you get a court to decide whether you have to or not.

Steven: OK. And do you have any experience with all, the whole NSA scandal that's been going on. I'm guessing even if you did you couldn't talk about, but have you been reading about it or following that news?

Rob: Yes.

Steven: If the NSA came to Civil Good, what would our options be? If they said we want access to your full database. The answer I've come up with so far is we don't really know because all of that is under gag order at the moment.

Rob: The answer is you would contact legal counsel right away to have a discussion about it. And depending on, a lot would depend on the nature of the request, and frankly what the request says. I suspect, the government, if the government comes to you and politely asks you for information you have the right to politely say no, the question is what's going to happen next. They could come back to you and say "OK now we've got a court order from some secret FISA court that requires this" or what, and that's on a case by case basis. I would not advise you to create any policies or procedures that address that issue one way or the other.

Steven: OK, and just leave it undetermined until the time comes basically?

Peter: Yeah, that's gonna be a really, that really will be a case by case scenario because there's this variety of different federal acts that could be invoked, Patriot act and how much wiggle room's involved in that, etcetera. That's the reason why a lot of European countries don't want to have United States cloud based services based on their soil because they're afraid that the Patriot Act is going to be able to reach right over there and circumvent any of their own state laws regarding privacy and yank the information out because the US company is subject to the Patriot Act.

Steven: Right, that's pretty much what I figured but thanks for the information. I think that's everything I had, is there any issue that you would recommend that I research further that I didn't bring up? Because I've brought up everything that I've found.

Peter: Ummm, just thinking here. I think you're good.

Rob: I think you want to have some community rules, I assume you already are going to but, about playing fair and not harassing people and maybe even warnings to people about exercising a degree of skepticism. I mean, I take it you're not doing anything to verify that this particular person has got an income over this amount and is a gun advocate and etcetera, etcetera, etcetera.

Steven: That's very true, the information is, we just ask for it and that's it.

Peter: Are the topics going to be open to any topic whatsoever?

Steven: That's the plan.

...

Peter: What Rob said is the less involvement you have with the conversations the better, I guess my question was whether or not, is this going to be restricted to people of 18 years of age or over?

Steven: That's a very good question, I hadn't thought of that. It's certainly a possibility but why would that be necessary?

Peter: It may not be, I'm just curious as to how this could morph into different areas that may or may not be suitable for minors.

Steven: That's a very good point. If we restrict it to people 18 years of age or older is that just putting a front page up that says you must be 18 years of age or older to enter and have a yes/no link or should we ask for age during sign up or...?

Peter: Well there are different ways you can deal with that, if you go onto a lot of alcohol sites that by law need to make sure that you're 18 years of age or over they just simply ask for a birth date verification which isn't very hard to circumvent quite frankly. But the, and this is going into targeting advertising which is different from what you're doing, but targeting advertising towards kids under COPPA they have a set of rules on independent verification and parental verification that wouldn't apply necessarily to you unless you're serving advertising to minor children but there are ways that you can have parental authentication there, but that may be going too far off field of what you're doing.

Steven: OK, that's definitely good to have, anything else you can think of?

Peter: I think you've pretty much covered it, I think Rob's point about community rules though is an important thing, that may not be necessarily something that's legal related, but it's something that will help you down the line if there's a legal challenge so that you can say to your community member who violated your terms and code of conduct that you can remove them or take whatever action you can, it gives you a leg to stand on.

...

## **E Privacy Policy**

### **Information We Collect from You**

- Civil Good collects information about your opinions and demographics. We use this information to match you with other users.
- We also use information about your opinions and demographics to sell targeted display advertisements. However, we do not allow third party advertisers to directly track your individual account. Clicking on an advertisement cannot be traced to your specific user account.
- We do not require that you provide any personally identifiable information. We collect your IP address for systems administration and security reasons, and we only store your email address if you sign up and provide it.
- We do not share information that can be directly tied to your specific user account with third parties. We will only comply with governmental organization requests for such information if they have followed the lawful procedures.
- We use your email address to send you verification links to reset your password. We also may use it to send you updates or sponsored advertisements. You can opt out of these updates or advertisement emails in your account settings.

### **Access**

- Any personal information you provide to us can be changed by you through your profile and settings page. You can disassociate your email address from

your account.

## Security

- All information you mark as private on Civil Good is encrypted until people with permission to access it do so. Information about your facts and opinions is not encrypted for performance reasons, but we make the best effort to protect your privacy.
- We follow generally accepted industry standards to encrypt private information submitted to us, both during transmission and after it has been received. No method of transmission over the Internet, or method of electronic storage, is 100 percent secure, however. Therefore, while Civil Good strives to protect encrypted information, it cannot ever guarantee its absolute security. If you have any questions regarding security on the Civil good website, please contact us at (needs phone number).

## Cookie Policy

- We do not store cookies, nor do we allow third party advertisers to use cookies.

## Redress

- Users who feel that their privacy has been compromised can report any issues to [privacy@civilgood.org](mailto:privacy@civilgood.org). We will do our best to resolve the issue within a reasonable amount of time.

## Updates

- Our privacy policy may change from time to time. If we make changes to this policy, we will notify you upon login. We are not required to notify you of changes to this privacy policy if you do not have a Civil Good account.
- We will also notify you via email of changes to the privacy policy if you have provided us with your email information.

## Consent

- This Privacy Policy is a part of Our Terms of Service. By visiting Our Site or using our services, you agree to be bound by the Terms of Service and this Privacy Policy.

## F Website Mockups

Mockup of original idea: <http://civilgood.wpi.edu/user/>

Mockup with recommended changes incorporated: <http://civil-good.wpi.edu/user/>